

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:36:50 ; Search time 6362.59 Seconds

(without alignments)
10782.626 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 ggttaagcgtcttactatg.....ttctctcagctatattta 1677

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: GenEmbl:

1: gb_ba:*
2: gb_neg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pi:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htg_hum:*
40: em_htg_mus:*
41: em_hgc_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1677	100.0	1677	AX509034	AX509034 Sequence
2	1677	100.0	91854	AC003680	AC003680 Arabidops
3	242.4	14.5	1483	AF085354	AF085354 Arabidops
4	102	6.1	1660	AY056214	AY056214 Arabidops
5	59.8	3.6	1141	AX083744	AX083744 Sequence
6	57.6	3.4	300695	AC079431	AC079431 Mus muscu
7	55.6	3.3	94348	AC109580	AC109580 Mus muscu
8	54.4	3.2	1058	AF015676	AF015676 Myrmecia pl
9	53.4	3.2	89551	AP004952	AP004952 Locust jap
10	53.2	3.2	151143	AC141740	AC141740 Apis mell
11	53.2	3.2	249943	AB014823	AB014823 Plasmodiu
12	52.6	3.1	56956	AC117081	AC117081 Dictyoste
13	52.4	3.1	2000	AX461144	AX461144 Sequence
14	52.4	3.1	94038	AC004401	AC004401 Arabidops
15	51.8	3.1	34750	AY217738	AY217738 Eimeria t
16	51.4	3.1	185747	EX511261	EX511261 Dario rer
17	51.4	3.1	349174	AB063522	AB063522 Mygilesw
18	51	3.0	65644	AL590364	AL590364 Human DNA
19	50.8	3.0	257109	AC115577	AC115577 Dictyoste
20	50.8	3.0	15635	AB083339	AB083339 Bombyx mo
21	50.8	3.0	15643	AF149768	AF149768 Bombyx mo
22	50.8	3.0	15656	AB070264	AB070264 Bombyx mo
23	50.8	3.0	15664	AF048187	AF048187 Bombyx mo
24	50.8	3.0	15928	AB070263	AB070263 Bombyx ma
25	50.6	3.0	8056	AX590446	AX590446 Sequence
26	50.6	3.0	183648	AC117076	AC117076 Dictyoste
27	50.4	3.0	7768	SVU28974	SVU28974 Spiroplasma
28	50.4	3.0	340552	PR929354	PR929354 Plasmodiu
29	50	3.0	810	CNS06EWO	AL395302 17 end of
30	50	3.0	969	DNA132902	AL122902 Drosophila
31	50	3.0	335050	PF929356	PF929356 Plasmodiu
32	49.8	3.0	244674	AL831726	AL831726 Dario rer
33	49.4	2.9	110000	PFMAL13_21	Continuation (22 o
34	49.4	2.9	110000	PFMAL13FP_2	Continuation (3 of
35	49.4	2.9	193187	AC116920	AC116920 Dictyoste
36	49.4	2.9	333321	AC116986	AC116986 Dictyoste
37	49.2	2.9	49843	AF288092	AF288092 Naegleria
38	49.2	2.9	160000	AB090308	AB090308 Bombyx mo
39	49.2	2.9	179484	AC010595	AC010595 Homo sapi
40	49.2	2.9	349980	AX344569	AX344569 Sequence
41	49	2.9	369	AX185123	AX185123 Sequence
42	49	2.9	64394	AL928596	AL928596 Human DNA
43	49	2.9	110000	PFMAL8P1_11	Continuation (12 o
44	49	2.9	349980	AX344558	AX344558 Sequence
45	48.8	2.9	653	AF140509	AF140509 Apis cera

ALIGNMENTS

RESULT 1
AX509034
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosoids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS
TITLE
Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
Stress-regulated genes of plants, transgenic plants containing

same, and methods of use
 Journal Patent: WO 0216655-A 3729, 28-FEB-2002;
 The Scripps Research Institute (US) ; Syngenta Participations AG
 (CH)

FEATURES
 source 1. 1677
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"

BASE COUNT 508 a 292 c 335 g 542 t
 ORIGIN

Query Match 100.0%; Score 1677; DB 6; Length 1677;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGTAAAGCGTTTACTTATGTTATATGCAACGGAAGATATTGCCATGTTGGAATGC 60
 1 GGTAAAGCGTTTACTTATGTTATATGCAACGGAAGATATTGCCATGTTGGAATGC 60
 61 TTTTCAGATGATCAAGGCTCCATAGATTTCTTGAAGGATGGTTTCAGGCTTTGTTA 120
 61 TTTTCAGATGATCAAGGCTCCATAGATTTCTTGAAGGATGGTTTCAGGCTTTGTTA 120
 121 GAAATTTGTTTATGCAACAGTATAGACATTAACATAGACAGATATGTAATGAAGA 180
 121 GAAATTTGTTTATGCAACAGTATAGACATTAACATAGACAGATATGTAATGAAGA 180
 181 TAAAGTTCTATGCTTAAAGAAATGACCCGATACGAATTAACAGATCATTAAGAT 240
 181 TAAAGTTCTATGCTTAAAGAAATGACCCGATACGAATTAACAGATCATTAAGAT 240
 241 TAAAGTTCTATGCTTAAAGAAATGACCCGATACGAATTAACAGATCATTAAGAT 300
 241 TAAAGTTCTATGCTTAAAGAAATGACCCGATACGAATTAACAGATCATTAAGAT 300
 301 AAAAACAATCGGAATCCAAACCTCAATTAACCAATACGACCAATTAATGATGCTGCG 360
 301 AAAAACAATCGGAATCCAAACCTCAATTAACCAATACGACCAATTAATGATGCTGCG 360
 361 TAAATGATGATATGCTATGATGAGGCAAAAGTGGTGGCTGCAAAATTAACAATATAC 420
 361 TAAATGATGATATGCTATGATGAGGCAAAAGTGGTGGCTGCAAAATTAACAATATAC 420
 421 TAAATGATGATATGCTATGATGAGGCAAAAGTGGTGGCTGCAAAATTAACAATATAC 480
 421 TAAATGATGATATGCTATGATGAGGCAAAAGTGGTGGCTGCAAAATTAACAATATAC 480
 481 AAAGCTAAAGATTAATCAACCGGTTGACCGGTTTAAATTTGGTAAATTCATATCC 540
 481 AAAGCTAAAGATTAATCAACCGGTTGACCGGTTTAAATTTGGTAAATTCATATCC 540
 541 AAAGCTAAAGATTAATCAACCGGTTGACCGGTTTAAATTTGGTAAATTCATATCC 600
 541 AAAGCTAAAGATTAATCAACCGGTTGACCGGTTTAAATTTGGTAAATTCATATCC 600
 601 GGTAAAGCGTTTACTTATGTTATATGCAACGGAAGATATTGCCATGTTGGAATGC 660
 601 GGTAAAGCGTTTACTTATGTTATATGCAACGGAAGATATTGCCATGTTGGAATGC 660
 661 GGTAAAGCGTTTACTTATGTTATATGCAACGGAAGATATTGCCATGTTGGAATGC 720
 661 GGTAAAGCGTTTACTTATGTTATATGCAACGGAAGATATTGCCATGTTGGAATGC 720
 721 TGTCTTACGCTCCCTCTTCTGCTGCTGCTATGATCAAGATGATAGTGGTTTAAAT 780
 721 TGTCTTACGCTCCCTCTTCTGCTGCTGCTATGATCAAGATGATAGTGGTTTAAAT 780
 781 CCCGAAACAAAGTACCAAGCAATCAAAATTAAGTTAAATCGGTTACATCTAGTACCGTGC 840
 781 CCCGAAACAAAGTACCAAGCAATCAAAATTAAGTTAAATCGGTTACATCTAGTACCGTGC 840
 841 AACTTACAATCAATTCGATTAATCTGATGATGATGATGATGATGATGATGATGAT 900

841 AACTTACAATCAATTCGATTAATCTGATGATGATGATGATGATGATGATGATGAT 900
 901 CGGATTTGATCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTAC 960
 901 CGGATTTGATCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTAC 960
 961 AAAGACGACGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1020
 961 AAAGACGACGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1020
 1021 ATTACCTTTTGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 1021 ATTACCTTTTGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 1081 GGGTTACAGCTTTTCCATTAATTAAGTAAATCTTTTGGCTTACCAATTAATTAATTAAT 1140
 1081 GGGTTACAGCTTTTCCATTAATTAAGTAAATCTTTTGGCTTACCAATTAATTAATTAAT 1140
 1141 ATTGAATATCTTCCATTAATTAAGTAAATCTTTTGGCTTACCAATTAATTAATTAATTAAT 1200
 1141 ATTGAATATCTTCCATTAATTAAGTAAATCTTTTGGCTTACCAATTAATTAATTAATTAAT 1200
 1201 AAAGCTAGATTTCAATTTGAGAGTATCTAGTATCTAGTATCTAGTATCTAGTATCTAGTAT 1260
 1201 AAAGCTAGATTTCAATTTGAGAGTATCTAGTATCTAGTATCTAGTATCTAGTATCTAGTAT 1260
 1261 TTCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1320
 1261 TTCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1320
 1321 AAAAAGAAAACATTTTGGTAAAGAAAGAAATTAAGTTTACGACCCCATTTGATGATG 1380
 1321 AAAAAGAAAACATTTTGGTAAAGAAAGAAATTAAGTTTACGACCCCATTTGATGATG 1380
 1381 TCCCATTAATTAATCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1440
 1381 TCCCATTAATTAATCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1440
 1441 TCGGAATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1500
 1441 TCGGAATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1500
 1501 AAAGCTGATTAATCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1560
 1501 AAAGCTGATTAATCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1560
 1561 AACTTAAAGTAAATCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1620
 1561 AACTTAAAGTAAATCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1620
 1621 GCTTTTAAACACTCGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1677
 1621 GCTTTTAAACACTCGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1677

RESULT 2
 AC003680/c 91854 bp DNA linear PLN 11-MAR-2002
 LOCUS Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence,
 DEFINITION complete sequence.
 AC003680
 VERSION AC003680.3 GI:20197048
 KEYWORDS HTG
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie
 1 (bases 1 to 91854)
 REFERENCE
 AUTHORS Lin X., Kaul S., Town C.D., Benito M.-I., Creasy T.H., Haas B.J.,
 Mu D., Walchli R., Rensing S.A., Koo H., Fujii C.Y., Utechtack T.R.,
 Barnstead M.E., Bowman C.L., White O., Nierman W.C. and Fraser J.C.M.

TITLE Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 91854)
AUTHORS Lin.X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
3 (bases 1 to 91854)
AUTHORS Town, C.D. and Kaul, S.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdctown@tigr.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6598396.
Address all correspondence to: atc@tigr.org

BAC clone F17K2 is from Arabidopsis thaliana chromosome 2 and is
near the molecular marker(s) FUS.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan* (Chris Burge,
<http://CCR-081.mt.edu/GENSCAN.html>), GenemarkHMM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/Genemark/>), GlimmerA (a variant
of GlimmerM, see Mihaela Pertea,
<http://www.tigr.org/softlab/glimmerm.htm>), and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
Simple repeats are identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

source
location/Qualifiers
1..91854
/organism="Arabidopsis thaliana"
/mol_type="Genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C1C02E07"
/clone="F17K2"
misc_feature
complement(1..1281)
/note="overlap with BAC clone F4L23 (AC002387:1..1281)."
57..747
/gene="At2g45480"
/note="synonym: F17K2.27"
join(<57..242,335..5747)
/gene="At2g45480"
join(57..242,335..747)
/gene="At2g45480"
/note="Protein sequence is in conflict with the conceptual
translation"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAFI8607.2"
/db_xref="GI:20197053"
/translation="DESKPSVQKFFPEVSDKCLEAKFSSNNRNDIARSREWNMMV
KNGLFHGFHPSDVLQGRGCPFLQGRCTDGGKWRCSKDVLSGGKYCD
KMHRSKKKHVPDITNSHENGAFSLTVAVRSVVPCKDQDKHVSVMGTLPR
VSDKSTSCSDTITITLALGDDDESYLSLSPGV"
complement(1002..2388)
/gene="At2g45490"
/note="synonym: F17K2.2; contains a protein kinase domain
profile (pDOC00100)"
complement(join(<1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..2388))

CDS
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"
/translation="MSKSTESDAGNTEKQWGLADFEIGRPLGKGFGRVYLAREAKS
KITVALKVFKEQIEKTHQIRREMEIQTSLRHPNLRFLFGWHDNERIELLEYA
HGGEYGVLSKNGHTTEOQAATVIASLSOALAYCHGKVIHRDIPKPNLLDHEGRLLK
IADFGSVLQSNKRTKTCQATLDYLAPEMVNDRDHDYAVDNWTGLICLYEFLVGNPFE
ASQKQTFKRLIKIDLSPLFTNVSEAKNLISQLVKDPSKRLSEIKIMQHPWIVKN
ADPKGVCASIDI"
complement(3414..5980)
/gene="At2g45500"
/note="synonym: F17K2.3; predicted by genscan"
complement(join(<3414..3616,3639..3825,3916..3965,
4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,
5506..5585,5670..5980))
/gene="At2g45500"
complement(join(3414..3616,3639..3825,3916..3965,
4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,
5506..5585,5670..5980))
/gene="At2g45500"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC06152.1"
/db_xref="GI:2979543"
/translation="MSFLRGIIDSFSSILNEBSKDPVSSSTSSSESMNGIDGVPVT
NERIAYLKGYPDLAKETAKGVRABEWGLHDDALLHYRNAQRINNEATSTSPSYIS
SSEKPVRSYRKISNMQVSEIRLQALGVSGENKRTVAVPSSASVSTASRYKRTL
SOKTPVARGVATERNPKDAAAPKPVKESGVNVDKLVEMINTIIVDSRSPVKDDV
AGLNGAKALLEMLIPAKRDLFTGLRRPARVTSILGLLFGPGNGKTMALAKAVAS
ESGATFNVSASSLTSKIDSIMSTRSENEASRLKSEFLIQDGVTSNPDDDLVII
IGATNPQBLDDAFLRLVRYIYFLPSPNVKLLFKLKQCPHSLDGDIDKIVKE
TEGKVLCTGCKHRTISQVTDYSGSDIQAACEAAMPIRELGANILTIQANKVLN
FSVQINVBVLSLKLGNRCVHVTKISTI"
6429..8288
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"
/translation="MELTSLIATVATTIFVLCTIYLMIRIFTGKSRNDRKVPVH
ATVEDLPHSDLYDETEIAREKETVRFSPQSEILTADPRNVHILKTRFDNYSK
GHSSRENMADLLGHGIFPAVDGKMWQKRLSPFSTVLRDPSCSVRRNASKLVGF
VSEFALSGRAFDADQLLMKTLDSIFKVGVELKCDGSEKGEFWEAFDEGNVAT
SRFIDPLWKLKFFNIGSQSKLSIATIDKFVYSLITTKRELAKQNTVREDIL
SRFLVSEKDPENMDKYLRIILNFMAGKDTAALLSWFLMLCKNPLQVETQVE
IRDVTPSHKTDVNGFVSEINEEALDEHLYLAALSETRLRYPVPVDMRCAENDV
LPHGVHSKGDNIYIAYAGRMVTYWGDAEEFKPERWLKGLFQPSPPKFIFFHA
GPRICLGKDPAYRQMKIVSMALLHFRFPKADENSKVYKMLTLHVDGLHLCAIPR
TSI"
complement(8412..9632)
/gene="At2g45520"
/note="synonym: F17K2.5; supported by full length cDNA:
Ceres.8220"
complement(join(<8412..8678,8768..8903,8987..9120,
9289..9356,9459..9632))
/gene="At2g45520"
complement(join(8553..8678,8768..8903,8987..9120,
9289..9356,9459..9612))
/gene="At2g45520"
/codon_start=1
/product="expressed protein"

/organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="2"
 /clone="RAF107-08-P04 (R10678)"
 /note="This clone is in a modified pBluescript vector
 (pLC-1) as a BamHI/XhoI insert.
 ecotype: Columbia"
 1..1660
 /gene="At2945680"
 1..1102
 /gene="At2945680"
 103..11173
 /gene="At2945680"
 /codon_start=1
 /evidence=experimental
 /product="putative PCF2 DNA binding protein"
 /protein_id="AA07063.1"
 /db_xref="GI:15810351"
 /translation="MATTQKLEEVAGKQDTLRAVDLITINGVRNVTSRFPQVNPVTS
 LEPKAEVPMPSPMSLAPSGTGPLKEASTKDRHTKVGRRIRMPATCAARI FOL
 TRELGHKSDGRTIRLLNAPPAIIAATGTCTVPAIAMSVMGT.LKIPPTNADSDGE
 NLMKKRKPSPNSSEVIDIDSAVSASSGLAPIATTITIQPOALASSVAQQLPQGNV
 PMWALPSNAPPTVGAFILIQIAGPSNPQLAFPAKASPSIYAAVQAQSTMARP
 PLYQVPSGFSVSDVSGSNLSRATSVNAPSSSGVTTGSSSIATTTTHLRDFSL
 EYERQELHQFMSITTAESSNH"
 1174..11660
 /gene="At2945680"
 464 a 400 c 385 g 411 t
 BASE COUNT
 ORIGIN
 Query Match 6.1%; Score 102; DB 8; Length 1660;
 Best Local Similarity 100.0%; Pred. No. 5e-11;
 Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1576 CTTCGACGTGAACCGCGTTAAACGATCTTCCCATGTATCCGCTTTTAAACATCT 1635
 Db 1 CTTCGACGTGAACCGCGTTAAACGATCTTCCCATGTATCCGCTTTTAAACATCT 60
 QY 1636 COTCGTCATCTCCACCGTCGCTTTCTCTCAGCTATATTTA 1677
 Db 61 COTCGTCATCTCCACCGTCGCTTTCTCTCAGCTATATTTA 102

RESULT 5
 AX083744/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 Kunat, L. and Clemens, S.
 Regulation of embryonic transcription in plants
 Patent: WO 011061-A 22 15-FEB-2001;
 UNIVERSITY OF BRITISH COLUMBIA (CA)
 Location/Qualifiers
 1..1141
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 1..1141
 /note="consensus sequence of A.t., L.a., and B.n. FAE1
 promoters"
 123 a 32 c 42 g 112 t 832 others

Promoter
 BASE COUNT 123 a 32 c 42 g 112 t 832 others
 ORIGIN
 Query Match 3.6%; Score 59.8; DB 6; Length 1141;
 Best Local Similarity 10.5%; Pred. No. 0.031;
 Matches 85; Conservative 299; Mismatches 422; Indels 3; Gaps 1;

QY 855 TCGATTACTTTCATCTGATTCCTAGTTCGGTTTGTATGTTTAATATCCGAGTGTACAA 914
 Db 1002 TTNAMWYATTRWAAAYAAAKWAERAGNNMRYGAAAGNKGCKAAAMATNGBWADTAGRVC 943
 QY 915 TACACAAGTACATAAGTAGTCGTATATGATGTATGTCGCGTTTAAATCAAGGACGACCAT 974
 Db 942 NNNNNWTTDVRPMAMKAKNNNNNNAYWTA CYNRAATNNKATWKMKTGHAHSKRTEH 883
 QY 975 AGGAGATTTTGGAAATCCCGAAAGAGGATTAATCCATAGACATTAATAGCTTTTGGT 1034
 Db 882 HTRTCERTKYNNNNNNARTVYVYHEAARNNMAMWTRTNNNNNNNNNNACNTRTWABW 823
 QY 1035 GCGCAGCGCTTGTGACCTACATTAATGGGTCCCAACCCCAAGTAGTGGCTTACAGCTTT 1094
 Db 822 KISWCNNNNNNNNNNNNNNWCHYTTANABCTRANNNNAARMACTNNYHAAVITTH 763
 QY 1095 TCATAAAATTAAGTAATCTTTTTCCTTACCAATAAAATTAATGAAAATCTT-- 1152
 Db 762 TDWCYKTMNTWYDMMTMBTTTTTRNMTTS TNNNNNNNNWACTNNNNNNMKAUYAH 703
 QY 1153 -TCCACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGTA 1211
 Db 702 ATNNWGCWNNNTDARTNNNTVMRRMMNTKTRWYSTTRHHYTGATNNNNNNNNNN 643
 QY 1212 TTTCAATTCGGAGTGTACTAGTAAGTACTAACCAGAAATGAGTTCTGATTTTG 1271
 Db 642 NNNNSCCTCTRMWTVRWTKGDMGTVRKVKWRDITCTYVDVWADSWWYANWMCREDV 583
 QY 1272 GATTTTGAAGCTTTCTTAGGTTAAAAAAGATATATTACTAAACAATAAAGAAAAAC 1331
 Db 582 TYTRNNNTYCKSYAHSYWYNNMAYRSYARNSSMARWTTTRNNMWSGBVRMRWAGTM 523
 QY 1332 ATTTTGTGAAAAGAGAAATAAGTTTACTGGACCCCATTTGTACAGATGCTCCATAATA 1391
 Db 522 WHEWNNNTDTRYVWWRWABTTTVDSCNAKSWRGNNWRMKNWMAANNNDAGM 463
 QY 1392 TACTGATAGAAGATAGACCAATGGAAAGTATTTGTTCACTGGTACAAATCGGAATGTT 1451
 Db 462 DHWTYMGNNNTMMERAWKMMNMAWCRRAYCCNNNNNACVWHKHKMWRWTKYMWKAAC 403
 QY 1452 CTTTAAAGCTCATCGAACACATCAGGACCGTTGATTTTCCCGCATCAAAAGCGTTGAA 1511
 Db 402 NNNNKAMTVRVAMWYSRDITNTDMMWTSDBWHWTYVDYTMWRANNNNNNNNRBCK 343
 QY 1512 TACTATTCCTCACTGTTTTCCTGCTCTATATATATATATCTCAGCAGTACACATTTAGTAA 1571
 Db 342 TTSMMWMDHMMTHCTYGNNTWGSAYBWAASMAAGASNBVTYVWCRWYMGKTWNN 283
 QY 1572 TCTCCTGGACGTGTAACGCGTTAAACGATCTTCTCCCATTTGATCCGCTTTTAAACA 1631
 Db 282 NNNKAWYRTKTVAVCNRRYYDPAVWTVTKRNYKYCYAYBYWYBYWYBYWYBYWYBYWY 223
 QY 1632 CTCTCGTCTCATCTCCACCGTCGCTTTT 1660
 Db 222 HRSNNWVKCNKYWVSHYHMYEYBK 194
 RESULT 6
 AC079431
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 Mus musculus chromosome 16 clone RP23-1J12, WORKING DRAFT SEQUENCE,
 77 unordered pieces.
 AC079431
 AC079431.1 GI:9958043
 HTG; HTGS PHASE1; HTGS DRAFT.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 300695)
 DOE Joint Genome Institute.
 Sequencing of Mouse

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 300695)
DOE Joint Genome Institute.
Direct Submission
Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 0
Center Clone Name: RPCI-23_1J12

Summary Statistics

Consensus quality: 164936 bases at least Q40
Consensus quality: 197223 bases at least Q30
Consensus quality: 214345 bases at least Q20
Estimated insert size: 257200; agarose-fp estimation
Quality coverage: 3.01 in Q20 bases; agarose-fp estimation
Quality coverage: 2.65 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 77 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1076: contig of 1076 bp in length
1077
1176: gap of unknown length
1177
2208: contig of 1032 bp in length
2209
2308: gap of unknown length
2309
3527: contig of 1219 bp in length
3528
3627: gap of unknown length
3628
4764: contig of 1137 bp in length
4765
4864: gap of unknown length
4865
5952: contig of 1088 bp in length
5953
6052: gap of unknown length
6053
7182: contig of 1130 bp in length
7183
7282: gap of unknown length
7283
8420: contig of 1138 bp in length
8421
9516: contig of 1096 bp in length
9517
9716: gap of unknown length
9717
11229: contig of 1413 bp in length
11230
11229: gap of unknown length
11230
12408: contig of 1179 bp in length
12409
12508: gap of unknown length
12509
13767: contig of 1259 bp in length
13768
13867: gap of unknown length
13868
15044: contig of 1177 bp in length
15045
15144: gap of unknown length
15145
16325: contig of 1181 bp in length
16326
16425: gap of unknown length
16426
17578: contig of 1153 bp in length
17579
17679: gap of unknown length
17679
18445: contig of 1167 bp in length
18446
18445: gap of unknown length
18446
20060: contig of 1115 bp in length
20061
20160: gap of unknown length
20161
21258: contig of 1098 bp in length
21259
21358: gap of unknown length
21359
22390: contig of 1032 bp in length
22391
22490: gap of unknown length
22491
23599: contig of 1109 bp in length
23600
23699: gap of unknown length
23700
24907: contig of 1208 bp in length
24908
25007: gap of unknown length
25008
26121: contig of 1114 bp in length
26221: gap of unknown length

26222
27273: contig of 1052 bp in length
27274
27373: gap of unknown length
27374
28472: contig of 1099 bp in length
28473
28572: gap of unknown length
28573
29583: contig of 1111 bp in length
29584
29783: gap of unknown length
29784
30334: contig of 1051 bp in length
30335
30835: gap of unknown length
30836
32035: contig of 1101 bp in length
32036
32135: gap of unknown length
32136
33334: contig of 1199 bp in length
33335
33434: gap of unknown length
33435
34598: contig of 1164 bp in length
34599
34698: gap of unknown length
34699
35818: contig of 1120 bp in length
35819
35818: gap of unknown length
35819
37082: contig of 1164 bp in length
37083
37182: gap of unknown length
37183
38445: contig of 1463 bp in length
38446
38445: gap of unknown length
38446
39768: contig of 1023 bp in length
39769
39868: gap of unknown length
39869
40973: contig of 1105 bp in length
40974
41073: gap of unknown length
41074
42174: contig of 1101 bp in length
42175
42274: gap of unknown length
42275
43341: contig of 1067 bp in length
43342
43441: gap of unknown length
43442
44532: contig of 1091 bp in length
44533
44632: gap of unknown length
44633
46068: contig of 1436 bp in length
46069
46168: gap of unknown length
46169
47421: contig of 1253 bp in length
47422
47521: gap of unknown length
47522
48629: contig of 1108 bp in length
48630
48729: gap of unknown length
48730
49986: contig of 1257 bp in length
49987
50086: gap of unknown length
50087
51394: contig of 1308 bp in length
51395
51494: gap of unknown length
51495
53158: contig of 1864 bp in length
53159
53258: gap of unknown length
53259
54802: contig of 1544 bp in length
54803
54902: gap of unknown length
54903
56009: contig of 1107 bp in length
56010
56109: gap of unknown length
56110
57452: contig of 1343 bp in length
57453
58667: contig of 1115 bp in length
58668
58767: gap of unknown length
58768
59224: contig of 1157 bp in length
59225
60024: gap of unknown length
60025
61159: contig of 1135 bp in length
61160
61259: gap of unknown length
61260
62764: contig of 1505 bp in length
62765
62864: gap of unknown length
62865
64293: contig of 1429 bp in length
64294
64393: gap of unknown length
64394
65793: contig of 1400 bp in length
65794
65893: gap of unknown length
65894
67168: contig of 1275 bp in length
67169
67268: gap of unknown length
67269
69245: contig of 1977 bp in length
69246
69345: gap of unknown length
69346
70850: contig of 1505 bp in length
70851
70850: gap of unknown length
70851
73220: contig of 2270 bp in length
73221
73220: gap of unknown length
73221
76908: contig of 3588 bp in length
76909
77008: gap of unknown length
77009
78773: contig of 1765 bp in length
78774
78873: gap of unknown length
78874
81789: contig of 2916 bp in length

* 81790 81889: gap of unknown length
* 81890 84399: contig of 2510 bp in length
* 84499: gap of unknown length
* 84500 87255: contig of 2756 bp in length
* 87256 87356: gap of unknown length
* 87357 91360: contig of 4005 bp in length
* 91361 91460: gap of unknown length
* 91461 95576: contig of 4116 bp in length
* 95577 95676: gap of unknown length
* 95677 99942: contig of 4265 bp in length
* 99942 100041: gap of unknown length
* 100041 105431: contig of 5390 bp in length
* 105431 105531: gap of unknown length
* 105531 112491: contig of 6959 bp in length
* 112491 121967: gap of unknown length
* 121967 122068: contig of 9377 bp in length
* 122068 132043: gap of unknown length
* 132043 132143: contig of 9975 bp in length
* 132143 146368: gap of unknown length
* 146368 156607: gap of unknown length
* 156607 166814: contig of 10108 bp in length
* 166814 177819: gap of unknown length
* 177819 179191: contig of 10905 bp in length
* 179191 191863: gap of unknown length
* 191863 191964: contig of 13944 bp in length
* 191964 204880: gap of unknown length
* 204880 220059: gap of unknown length
* 220059 220159: contig of 15079 bp in length
* 220159 238639: gap of unknown length
* 238639 238739: contig of 18480 bp in length
* 238739 262566: gap of unknown length
* 262566 262667: contig of 23827 bp in length
* 262667 300695: gap of unknown length
* 300695 38029: contig of 38029 bp in length.

FEATURES
source
Location/Qualifiers
1..300695
/organism="Mus musculus"

Query Match 3.4%; Score 57.6; DB 2; Length 300695;
Best Local Similarity 53.1%; Pred. No. 0.038;
Matches 120; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 1133 TAAATATTTGAAATCTTCCACCATGAGAAAGTTAATTTGATCGGATGGAAT 1192
Db 37687 TAAATGTAATTTGAAAGAAAGTTGTTAAATTAAGGAAATTTATAGTAAATTAAT 37746
Qy 1193 TTTTGTCAAGCTAGGTATTTTCAATTTGGGAGTGTACTAGTAACTAGTAACTAACCA 1252
Db 37747 TATTAATGAGGNTATTTATTTAAATGGGAGTTTAAATAAAATAAAGTAAGAGAT 37806
Qy 1253 CAATGAGTTCTGATTTGGATTTTGAAGCTTTTCTAGGTAAATAAACAAGTATATTAC 1312
Db 37807 ATTAGAATAAGAAATATTAAATAGTTTATATAAAATATAAATAATTAATATAGTAAATTTA 37866
Qy 1313 TAAACAAATAAAGAAAGAAACATTTTGTGAAAGAGAAATAAGTTTA 1358
Db 37867 AAATATAGAGAAAGATTATTAATTAAGAGAAAGAAATATTTGTAA 37912

RESULT 7
AC109580 94348 bp DNA linear VRT 19-NOV-2002
LOCUS
DEFINITION
Dario rerio clone busml-183j13 strain AB, complete sequence.
ACCESSION
AC109580.14 GI:24635954
VERSION
AC109580.14 GI:24635954
KEYWORDS
HTG.
Dario rerio (zebrafish)
Dario rerio
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 94348)
Kupfer, D., So, S., Wang, H. and Roe, B.A.
Dario rerio PAC Clone busml-183j13
Unpublished
2 (bases 1 to 94348)
Kupfer, D., So, S., Wang, H. and Roe, B.A.
Direct Submission
Submitted (05-FEB-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 94348)
Kupfer, D., So, S., Wang, H. and Roe, B.A.
Direct Submission
Submitted (01-NOV-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 94348)
Kupfer, D., So, S., Wang, H. and Roe, B.A.
Direct Submission
Submitted (06-NOV-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 94348)
Kupfer, D., So, S., Wang, H. and Roe, B.A.
Direct Submission
Submitted (19-NOV-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Nov 6, 2002 this sequence version replaced gi:24462347.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code: UCKNOR

Location/Qualifiers
1..94348
/organism="Dario rerio"
/mol_type="genomic DNA"
/strain="AB"
/db_xref="taxon:7955"
/clone="busml-183j13"
/clone.lib="Zebrafish PAC library"
BASE COUNT 305030a 17974 c 16523 g 29348 t
ORIGIN

Query Match 3.3%; Score 55.6; DB 5; Length 94348;
Best Local Similarity 47.9%; Pred. No. 0.12; 174; Indels 0; Gaps 0;
Matches 160; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

Qy 1092 TTTTCCATAAAATTAAGTAAATCTTTTTCCTACCAATAAATTTATTTGAAATCT 1151
Db 87900 TTAACATATAATAAATTTGTAAATAAACTACAAATTAATTCATATAATCAATGAA 87959
Qy 1152 TTCACCATAGAAAAGTTAAATTTGATCGAGCGATGGAATTTTGTACAACTAGTA 1211
Db 87960 ATAAATGAAATTAATAAAGTTAAATTAATAATATAATTTATTTAAAGCAATAA 88019
Qy 1212 TTTCAATTTGGGAGTGTACTAGTAACTAGTAACTAACCAATGAGTTCTCTGATTTG 1271
Db 88020 TCTATTTTAAATGAACTAAATTTTAAATTAATCAATCAATCAATCAATTTAA 88079
Qy 1272 CATTTTGAAGCTTTTCTAGGTTAAATAACAGTATATTACTAAACATATAAGAAAC 1331
Db 88080 AAAAATGCAAAATTAATTCATTTAAATAAAATAAATTAAGTAATTAATATATAAC 88139
Qy 1332 ATTTTGTGAAAGAGAAATAGTTTACTGGACCCCATTTGTACAGATGGTCCCATATAA 1391
Db 88140 TCCATTTATTTAAAGCAATATTTTACATTTATTTTATTTAAATTAATTAATTA 88199
Qy 1392 TACTGATAGAGATAGAGCAATGCAATGCAATGATTT 1425


```

Db      88200 TATTATTAAATGCAATAAATAAATTAAATTAATTT 88233

RESULT 8
LOCUS   MP15676 1058 bp DNA linear INV 09-FEB-1996
DEFINITION Myrmecia pilosula H187-128 mitochondrial cytochrome b gene, partial cds.
ACCESSION U15676
VERSION   U15676.1 GI:576754
KEYWORDS  mitochondrial Myrmecia pilosula (Australian jumper ant)
SOURCE   Myrmecia pilosula
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Formicidae; Myrmecinae; Myrmecini; Myrmecia.
REFERENCE 1 (bases 1 to 1058)
AUTHORS  Crozier,R.H., Dobric,N., Imai,H.T., Graur,D., Cornuet,J.M. and Taylor,R.W.
TITLE     Mitochondrial-DNA sequence evidence on the phylogeny of Australian Jack-jumper ants of the Myrmecia pilosula complex
JOURNAL   Mol. Phylogenet. Evol. 4 (1), 20-30 (1995)
MEDLINE   95346030
PUBMED    7620633
REFERENCE 2 (bases 1 to 1058)
AUTHORS  Crozier,R.H.
DIRECT SUBMISSION
TITLE     Submitted (08-OCT-1994) Ross H. Crozier, School of Genetics, La Trobe University, Bundoora, Victoria 3083, Australia
JOURNAL
FEATURES
source    Location/Qualifiers
1..1058
/organism="Myrmecia pilosula"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/strain="H187-128"
/db_xref="taxon:13618"
/issue_type="head and thorax muscle"
/dev_stage="adult worker"
<2..733
/codon_start=1
/transl_table=5
/product="cytochrome b"
/protein_id="AA86970.1"
/db_xref="GI:576755"
/translation="FMGATVNLVSTIPFLGSDITQWLWGYSVSNATLGRFSIHF
ILPFMLIMVIMHLYPLHUTGNSPLGINSNLKIPENLYSTKDLGLFTTLFLFTL
LIIMTVLPSDPDNPFLPANPLVTPHIOPEWLEFAYSLISIPKLGVALPSSIC
ILYFELMNSNSSTFPLSQLLINLFMTFMILTGLSQVIEHPYIIMAQFFSMLY
FNFILKFSYKXWDYFIYNKFNN"
misc_feature 734..1058
/notes="intergenic space"
BASE COUNT 398 a 132 c 61 g 467 t
ORIGIN
Query Match 3.2%; Score 54.4; DB 3; Length 1058;
Best Local Similarity 48.1%; Pred. No. 0.42;
Matches 154; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

2Y 1092 TTTTCCATAAAATTAAGTAAATCTTTTTCCTAACCAATAAAATTTATTGAAATCT 1151
|||
2b 739 TTTTATTAAATAATTTATTTAATAATATTTATTATAAATAATTTATTATAAATAA 798
|||
2Y 1152 TTCCAACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGTA 1211
|||
2b 799 TTTTATTGATTAAAAATTAATTAATAATTAATAAATAATTTTATAAATAATTTAAAT 858
|||
2Y 1212 TTTTCATTTCGGAGTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGT 1271
|||
2b 859 ATTATAAATTTATTATTATTATTATAATAATAATAATAATAATAATAATAATAATA 918
|||
2Y 1272 GATTTTGAAGCTTTTCTTAGGTTAAAAAACAAGTATATTACTTAAACAATAAAGAAAAAC 1331
|||
2b 919 TAAATTTATATATAATTAATATAACAATAAATAATTTATTATATTATATAAATAATTA 978
|||

```

```

QY 1332 ATTTTGTGAAAGAGAAATAAAGTTTACTGGAACCCATTTGTACAGATGGTCCCATATAA 1391
|||
Db 979 ATTATATTAATTTCTATATCTAACTAACTAACTAACTAACTAACTAACTAACTAACTAA 1038
|||
QY 1392 TACTGATAGAGATAGAGCA 1411
|||
Db 1039 AAAATATTTAAGTTAATGAA 1058
|||

RESULT 9
LOCUS   AP004952/2 89551 bp DNA linear PLN 19-JUL-2002
DEFINITION Lotus japonicus genomic DNA, chromosome 1, clone: LJ19C08, TM0125, complete sequence.
ACCESSION AP004952
VERSION   AP004952.1 GI:21907970
KEYWORDS  HTG.
SOURCE   Lotus japonicus
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
REFERENCE 1
AUTHORS  Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T., Sato,S. and Tabata,S.
TITLE     Structural Analysis of a Lotus japonicus Genome. I. Sequence Features and Mapping of Sixty-six TAC clones which cover the 6.7 Mb Regions of the Genome
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 89551)
AUTHORS  Nakamura,Y.
DIRECT SUBMISSION
TITLE     Submitted (26-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynk@kazusa.or.jp. URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935, Fax: 81-438-52-3934)
JOURNAL
FEATURES
source    Location/Qualifiers
1..89551
/organism="Lotus japonicus"
/mol_type="genomic DNA"
/db_xref="taxon:34305"
/chromosome="1"
/clone_lib="LJT library"
/notes="TAC clone: TM0125"
BASE COUNT 27663 a 17797 c 17022 g 27069 t
ORIGIN
Query Match 3.2%; Score 53.4; DB 8; Length 89551;
Best Local Similarity 45.6%; Pred. No. 0.34;
Matches 224; Conservative 0; Mismatches 266; Indels 1; Gaps 1;

QY 1055 ATTAATGGGTCCAAACCCAGATATGGGCTTACAGCTTTTCCATAAAATTAAGTAAT 1114
|||
Db 73682 ACTAATTTGGAATGACTGAAAAGTAACTAAATAATTTATCCTTTAAATAATAGTTT 73623
|||
QY 1115 CTTTTTCCTCAACCAATAAAATTTATGAAATCTTTCCACCACATAGAAAGTTAAAT 1174
|||
Db 73622 CATATTTTAATAATTTATTATTATTGAAATAAATCAATAATGATACAAAGATACAA 73563
|||
QY 1175 TTGATCAGCGATGGAATTTTGTACAAAGCTAGGTATTTCATTTTGGGAGTGTACTAGTA 1234
|||
Db 73562 ATGATAATTTATTATTATTTACAAATAATGTAATAATTTCCCTTTAAATATATTACTGTA 73503
|||
QY 1235 ACTAGTAAGTACTAACCCAGATGAGTTTCTGATTTTGGATTTTGAAGCTTTTCTAGGTT 1294
|||
Db 73502 AATAGACGAAATAAAAATAATTAATTAATAGATTTTGTATAAAGTATT -TTTTTG 73444
|||
QY 1295 AAAAAACAAGTATATTACTTAAACAATAAAGAAAAACATTTTGTGAAAGAGAAATAAG 1354
|||
Db 73443 AAAAAATGATAAATAATTTACTTCTATAAATAAAGTATTATTTTTAAATAATAATAA 73384
|||

```

QY 1355 TTTACTGGACCCCATTTGACAGATGGTCCCATATAATATCTAGTAGAGATAGCAATG 1414
 |||||
 Db 73383 TTTACATTTTAAATTTAAAGTAGACTGTGAATAAGCTTAAATAAGGAATATCTCTA 73324
 QY 1415 GAAAGTGTATTTGTTTCCACGTGGTCAATCGGAATGGTCTTTTAAAGCTCAACGACACATC 1474
 |||||
 Db 73323 GCTATATGCTAGCGTGAATTTATCATCATGGTATGATTAATTTCTTTATTTGGGAATATTT 73264
 QY 1475 AGGACCGTTGATTTTCCCGCATCAAAAGCGTTGAATCTATCTCTACATGTTTTCCTG 1534
 |||||
 Db 73263 CAATTGGTAACCTTATCTCTTTTGGTAATAAATTAATTAATTTCTCTATAAATTTTTA 73204
 QY 1535 CTCCTATATAT 1545
 |||||
 Db 73203 CTCCTATATAT 73193
 |||||
 RESULT 10
 AC141740
 LOCUS
 DEFINITION Apis mellifera clone CH224-58F11, WORKING DRAFT SEQUENCE, 10
 AC141740
 AC141740.1 GI:29123924
 HTG: HTGS PHASE1; HTGS DRAFT.
 Apis mellifera (honeybee)
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Hymenoptera; Apoidea; Aculeata; Apoidea;
 Apidae; Apis.
 1 (bases 1 to 151143)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbacia,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,B., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Dreher,H., Dugan-Rocha,S., Durbin,K.J.,
 Earhart,C., Edgar,D., Edwards,C.C., Eihaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulesed,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oghu,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojebokan,I., Roife,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
 Sodergren,S., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Umanji,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,O.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs.R.
 Direct Submission

TITLE

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 151143)
 Worley,K.C.
 Direct Submission
 Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center of Medicine
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: AMEK
 Center clone name: CH224-58F11
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 148444 bases at least Q40
 Consensus quality: 149298 bases at least Q30
 Consensus quality: 149814 bases at least Q20
 Estimated insert size: 147256; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 10 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

* 1 1058: contig of 1058 bp in length
 * 1059 1158: gap of unknown length
 * 1159 2581: contig of 1423 bp in length
 * 2582 2681: gap of unknown length
 * 2682 4880: contig of 2199 bp in length
 * 4881 9305: gap of unknown length
 * 9306 9405: contig of 4325 bp in length
 * 9406 12997: gap of unknown length
 * 12998 13097: gap of unknown length
 * 13098 22656: contig of 9559 bp in length
 * 22657 22756: gap of unknown length
 * 22757 36494: contig of 13738 bp in length
 * 36495 38595: gap of unknown length
 * 38596 53261: contig of 18666 bp in length
 * 53261 53361: gap of unknown length
 * 53361 96148: contig of 42787 bp in length
 * 96148 96248: gap of unknown length
 * 96248 151143: contig of 54896 bp in length.

FEATURES
 source

1..151143
 /organism="Apis mellifera"
 /mol_type="genomic DNA"
 /db_xref="taxon:7460"
 /clone="CH224-58F11"

BASE COUNT 47501 a 27383 c 27767 g 47591 t 901 others
 ORIGIN

Query Match 3.2%; Score 53.2; DB 2; Length 151143;
 Best Local Similarity 45.8%; Pred. No. 0.35;
 Matches 184; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 1027 TTTTGGGGGGCGACCTGTGACCTACATTAATGGGTCCACCCCAAGTATGGGCTTA 1086
 |||||
 Db 125556 TTTTGGGAAGAGATATTTCTTAATTTGAAGATGGCCCAAGTCAATCTGCTTTA 125615
 QY 1087 CAGCTTTTCCATAAATAAAGTAATATCTTTTTCCTTAACCAATAAATATTGAA 1146
 |||||

YTNISENLVYRKISITETSSDRSSSDNVLLENINNVLVKYVSGLDYTLNIRISXDK
 IBEIKFYIITHNNIQLANDYSTLDEVINILIKOKONIKLOKMLFIIKKNVDFYK
 NSEIYFYPYCNVLTIKYCNDRNCELTYYTDNDEKNS"TENSYLSSNNMYNT
 NYSNHSNLSIKADTSSNKNCSPIFGINERNSPNFGINESIKWKDKTOINNDYVLI
 MRYNLNESHVILIKDRVTSNPNKISKMDNRYNSPNFIEHLHNNDKTOINNDYVLI
 NNWDGKNLNNSTDFYOSISDFYKNEKNLICOENNNYTIKCAFFLNSNETHV
 LTNKNNHSGKNLISLSCDFDELSALTINDSSLNKAIYEFNNSKOLVEDICNT
 NYNDINHEMYTDRDERKMDDEYCNGDYNILFRSKQIKKCKELLSCISYNEKENYN
 DQDMNINLNDTKVNLITDGNISPDENIGINNHYNTINLKVITYEANKSEYGNNE
 RLMPITKQIILANKYKVLVSKTQSTTLKCLNLLIKVKVTDIQTVPFLPYCHYKMDDS
 EITHDKNKQDKFVNLNTIKKKNENYRQHDIKNLDHKNHQIINNKKBVPBKVC
 LKWKYKNGKQFLOGGLELWVLNLCNANTNNNLSNKNILQIYDSFYVKEHLIIVTEVM
 QSDLYNFRKKGKLTGLQOLIKLTKLLEGLAYHISNLIHCDLKBPENIMNKKKX
 NHEKGNKNGVNSVNIYNDTIEPNNKMLDNNILYCNKIKNDNNNNNNNNNNPHN
 QYNNKSTNVLVDSGNSNPNKMDNNILYCNKIKNDNNNNNNNNNNNNPHN
 INTYNTQKDKIKIDIPNSCIYESDKEMVLOTQSYRESPEVLQONQYDKRIDITWSLGC
 ILPEFTIKLIPQYQINRYFVYSIGYPPFPFYMNNCRPIPIFTKHGLIILKCTT
 DNMYENIKEOLOQDDEBEIVSNSNDFRLNKNLNTIKOLLKKNKPNSTSPKRN
 NNNEIYDVVCYPSDNLKKNFQISDTLFDVDFLSLLQIDPSKRCNAMEALKHPWLQPN
 LYKGL"

1089. 1149
 /rpt_type=tandem
 4186. 4225
 /rpt_type=tandem
 4358. 4387
 /rpt_type=tandem
 4543. 4577
 /rpt_type=tandem
 5071. 5152
 /rpt_type=tandem
 5352. 5382
 /rpt_type=tandem
 complement (-5667. ->7391)
 /locus_tag="PF14_0409"
 complement (-5667. ->7391)
 /locus_tag="PF14_0409"
 complement (5667. 7391)
 /locus_tag="PF14_0409"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA397022.1"
 /db_xref="GI:23497480"
 translation="MFNVFSAFLDNNINKNDNVLGKTLTNANNERTKNLPLK
 RLATVANGVONLEKKNINKNTDNEBKTANNFKTNGIENTPTTINESTFRKNY
 SVVNVKONLEKKLKRKKRKHVILMERKQKREKRQKQKMDQDQEQYDDE
 EHEEGEDDNDDEYDQKNEODSEHDEEDYYAENDHNNNNNNNNINRSFKNRSS
 KQKYDENDFNHKKQKGDDEYDDDEDDDEDDDEYDGGGYSKYNNNSYDNTFNKN
 YHSGNPNFHTTSSNNHNNNSQKQKFKRKKKCKEYERARLKEKHQK
 QNSLGSILNKTSFNKNDNSTQBEATVKYRRGRGPKASEVAAINWNKKLQNDIRKY
 YKSNQSGFMKCNWNEKNSQBEANTVFKIIEKSIQENSINFMKKKPNFIVLEK
 NYKNNIKGKVTICPLNYQSKGAMAILIGDTITYPGVKNSYGFMTFLVLDCHTNFF
 IDTGINKNVYICEKEMQLLISFGDMYFKQSHMEMEARLILLVCNKNQPFQVQKHIX
 DPMQNNK"
 7729. 7782
 /rpt_type=tandem
 7979. 8020
 /rpt_type=tandem
 8386. 8438
 /rpt_type=tandem
 9083. 9109
 /rpt_type=tandem
 9264. 9293
 /rpt_type=tandem
 <9317. ->11238
 /locus_tag="PF14_0410"
 join(<9317. 9689.9836. ->11238)
 /locus_tag="PF14_0410"
 join(9317. 9689.9836. 11238)
 /locus_tag="PF14_0410"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA397023.1"
 /db_xref="GI:23497481"


```
/translation="MRKLLSTALMDLKKIKNNPFESKXNDNLSKLIIRNIHKPFENDE
ILKIERKNEYKDTKILKCSYDFVFNKVRYSFDQINKIIRLYNISEIYDIHFNTS
VFNVIIRKLDIAIPYVWVFNHRLDYNNINILKEVEIKKIDFNNDLDTIILSFTI
LOELLKKIIPVNNKQVSENIYDTKDHSLCSEALYDMDIFKILIKIOKQKNIHEN
LSVNSJLLNMLSRINFENYVEIEKFTKYNKYNLNDKDLPEHFPILLNSPAKCNH
INIMKYLKYNKNGNFENNI-SYVNITNAVHYMAKFNRYNATFLNHLKOKVIBLIDLP
QRESFNMSLAKUHIKDDFYIYAFQIKQIDVMMMSVAQIOLDMRRKKNVSGQ
DIGATIEKKLENDIISCPYDIKDTLNGONIDYKNNIEHNKYNHNEEKVKEKEIIT
TQTTQNKDNLNNTQLTQPPENNIISEKNTINTYIISDLEKNIHLHLVKNYIKHI
QCHSLHLVTOVFFCQLQNTINSIDYVYKSLERLKRKNDMTLNLIVAKYFLRIFIEK
QEAHFQKLPRLSKQFAKEILLSDNN"
9716..9811
/rpt_type=tandem
repeat_region
11420..11447
/rpt_type=tandem
repeat_region
11803..11849
/rpt_type=tandem
repeat_region
11893..11918
/rpt_type=tandem
repeat_region
<12051..>12703
/locus_tag="PF14_0411"
gene
join(<12051..12111,12315..12394,12539..>12703)
mRNA
/locus_tag="PF14_0411"
CDS
join(12051..12111,12315..12394,12539..12703)
/locus_tag="PF14_0411"
/codon_start=1
/product="small nuclear ribonuclear protein, putative"
/protein_id="AAN37024.1"
/db_xref="GI:23497482"
/translation="MATISGETFLPLALMDKICIGSKIWMKGDKEIVKGLVGFDEY
VNVLEDEYETVYANNVKNKIKKLLNLGLNITIMVPGTVPVNYDYBEKLEBENIA"
12495..12537
/rpt_type=tandem
repeat_region
complement(<12999..>15647)
gene
/locus_tag="PF14_0412"
mRNA
complement(join(<12999..15334,15524..>15647))
/locus_tag="PF14_0412"
CDS
complement(join(12999..15334,15524..15647))
/locus_tag="PF14_0412"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAN37025.1"
/db_xref="GI:23497483"
/translation="MNPDCDTRKFRDISEKIHGKYVFIIPKGGSKYIKNDYKNEADI
NNISYASEFLNNVDIYNDKYGINKVCFSPQKIVAGCTSGNFIVLYDLYENKLLSK
ICTRIDIQIDAIISDNQKIVACENRLIEIFDIENRKRKNMDIGKVDVSIPIII
KNTLKNKNCSESDNKHNNIILSKNLDNKIIPYHKTKYKINSKGSIVIPLYNYED
LIKSNIMDLNLYLKNKNEIINKSFELIKESHEVITSCIAIPNCFNLVYSGGDG
VLKWDLRMLNFIKSKSSNSKSDIIFDNKPPASICSHEDVTISITFNKIIDHEG
SYLKKQKQKKKIFNSKSCIDNKNVYSSLPDENNDSTSDYEDSDNSVSSP
HSSICYSGLTFDLSKDFENILMTSGYDGYRILYDINSNLIKSFYDEBSIITHCFSN
NKKYIISITNKSFKQKDFVFMNKNMKNMAYLANNYKTYCLNKPLIKENLENNY
HDTINEDDIFSSQYNDKVALCKIGNIFDEBLEERI KIVNELNKRINEQENKPEP
NKNDNNNNNNNNNNKNDKKNKKNYKKNYNNNDNNITVEHDDYVPSNYS
NSDSIYDENTIKSTYEVHVNKLEPTWLSLFDVNLKYMDLPYDNDINSKYNVY
KSYFNHNNFLNTHNSLNDKPELSYMLSEIISNEDDIPFSSCIAGDKCIIPIISD
TYAHVYDLYTGFGVHNSIKNLYPNYVDNHSYSHSIDTNKKRISFLTSDVTPKPNQ
IIATNSGYPGDSIVLWAFAR"
13628..13761
/rpt_type=tandem
repeat_region
15450..15497
/rpt_type=tandem
repeat_region
15809..15855
/rpt_type=tandem
repeat_region
complement(<17033..>19621)
gene
/locus_tag="PF14_0413"
mRNA
complement(<17033..>19621)
/locus_tag="PF14_0413"
CDS
complement(17033..19621)
/locus_tag="PF14_0413"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAN37026.1"
```

```
/db_xref="GI:23497484"
/translation="MSSVFNFYFKNLFVKKRNLFFKYNILERYSKITSVNLNNMN
IHKHIIQKINNSDFVSDIVEVTGLHLDXDRVSLDSVEAHCVCAKSPFCQIGISIA
KKKDIVELYOTDENNKIKKKS SVHVRKVIYENGNNKICNIEKKQEHISPYCIY
VPPENKYPVSSTSLSPLENNEPDMNEMFNGVGLRPNNEBKKNIVKIKNELNN
LLSKNTNNKNNIIFGEKIHDKIQIETENYKIVDDEKEALISIIKKIKWINNTND
Query Match 3.2%; Score 53.2; DB 3; Length 249943;
Best Local Similarity 43.4%; Pred. No. 0.32;
Matches 294; Conservative 0; Mismatches 383; Indels 1; Gaps 1;
QY 684 ACATTCACATTTTCGTTTGTAGCCTTTCTTTCTGGTGTCTAGCGTCCCTCTTCCTGGT 743
Db 190305 ATATACTCCTGTGCATTTATATAAAAAATATGTTGAATACATAAAATATCTTTATCTACA 190246
QY 744 CGTCTGTATGTACAAGTAGCATAGCTAGTGTGTTCCAAACCCGGAACAAGTAGTCCACGAATC 803
Db 190245 CATATAATCTTCACATTTAATAATAAAAAACAAAAACAAAAACAAAAACAAAAACAA 190186
QY 804 AAAATAAGTTTGAATCGGTACATCTAGTTACCGTCCGACCTTACAACTTCAATTCATTCGATTACT 863
Db 190185 AAAAAACATAATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 190126
QY 864 TTGATCTCATTTCTAGTTTCGGTGTGTTTGTATGTTTAATATCCGATTTGTCAAGTAGTACACAAGT 923
Db 190125 ATGTTTTTTTTTATTTATTTATCTATTTTTTAATATATATATATATATATATATATATAGGT 190066
QY 924 ACATAAGTATCGGTATATGTATGTACCGGTTTAAATCAAGGACGGACGATAGGAGGATT 983
Db 190065 ATTTGAATTTATATATTTCTTTTGAAGAATTAATAAAAAACAAAAACAAAAACAAAAACAA 190006
QY 984 TTGGATCTCGAAGAGGATTATTTCCATGACACTAATTAGCTTTTTTGGTGGCGCAGCC 1043
Db 190005 TATATATATATATATATATATATATATATATATATATATATATATATATATATATTTT 189946
QY 1044 TTGTGACCTACATTAATGGGGTCCAAACCCCAAGTAGTGGGTTACAGCTTTTTCATAAAA 1103
Db 189945 GACAAAACATATATTTAATTAATTCATATATTTTGTAAACCTGTTATAATTAATTAACGATAAAC 189886
QY 1104 TTAAGTAAATCTTTTTTGGCTA-ACCAATAAAATATTGAAATCTTTTCCAAACCAT 1162
Db 189885 AAAATATATCAATATGTTTATATATATATATATATATATATATATATATATATATATATTTT 189826
QY 1163 GAAAGTTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGTATTTTCATTGGG 1222
Db 189825 ATTTATTAATTTGTTTGAATTTATTTTAACTCTTAATAGTAACACGTCACAATCTTT 189766
QY 1223 AGTGACTAGTAACCTAGTAAGTACTAACCAAGTAGGTTCTGATTTTGGATTTTGAGC 1282
Db 189765 TTTATATATTTTATATATATATATATATATATATATATATATATATATATATTTTTTTTAA 189706
QY 1283 TTTTCTTAGGTTAAAAACCAAGTATTTTACTAAACAATAAAAGAAAAACATTTTGTGAAA 1342
Db 189705 CTTTATTTTATTTATATATATATATATATATATATATATATATATATATATATTTTATCGAAT 189646
QY 1343 AGAGAAATAAGTTTACT 1360
Db 189645 TAAAAATTAATGTTTATT 189628
RESULT 12
AC117081
LOCUS
DEFINITION Dictyostelium discoideum chromosome 2 map 4903137-4900091 strain
AX4, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC117081
VERSION AC117081.1 GI:20066271
KEYWORDS HTG; HTGS_PHASE2
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
AUTHORS Gloeckner,G., Eichinger,L., Szatranski,K., Pachebat,J., Dear,P.,
```

Lehmann, R., Baumgart, C., Farra, G., April, J. F., Guigo, R., Kumpf, K.,
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
Noegel, A. A.
Sequence and Analysis of Chromosome 2 of Dictyostelium
The Dictyostelium Genome Sequencing Consortium
2 (bases 1 to 56956)
Baumgart, C.
Direct Submission
Submitted (08-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
3 (bases 1 to 56956)
Baumgart, C.
Direct Submission
Submitted (24-MAY-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
CDS predictions from GenesID may contain errors. Further Information
is available from IMB Jena, Department of Genome Analysis
(<http://genome.imb-jena.de/dictyostelium/>)
and the University Cologne, Institute for Biochemistry I
(<http://www.uni-koeln.de/dictyostelium/project.shtml>)
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..56956
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
map="4303137-4960091"
join(33..523,639..2638)
/note="ORF_ID:dd_02927"
/pseudo
/codon_start=1
join(3306..3316,3871..4174)
/note="ORF_ID:dd_02929"
/product="hypothetical protein"
/protein_id="AAM43758.1"
/db_xref="GI:21166141"
/translation="MIVCILSTSFSSKTHSTAFENDLKDSEIWEVNIISLMLP
KFNKVGQVQLEIEGYNQSWNSIKIGNVTNIYIRTNLTFFPDHLSNIRSDSIE
KN"
Join(4284..4740,4840..5898,6096..6795,8057..8351,
8472..8483)
/note="ORF_ID:dd_02930"
/pseudo
/codon_start=1
join(8354..8906,9008..11010)
/note="ORF_ID:dd_02932"
/product="hypothetical protein"
/protein_id="AAM43759.1"
/db_xref="GI:21166142"
/translation="MNSGFPSSNRTTRKLIKIDISNITGDTSDSCQLIIVTN
NSMGLPTCTCFPEINLERYFIGNRFINLNTSSPSIIPNIVFNNTSIEGS
IVFYELIFGNLGGSWFFFAVSDSVGTMIEMIPNSLFLKFCNGEFPFLNLSY
SKSNPTFTLUGLIPILNLDVQSWTDLVIDGSPFTYNSNTINTEYNCNLVTN
TNFYIKTLENSIDQSLIEDIISVTNGNLTQTLNPNVINTLNTENYDCNGN
YCLSDIGKCDPTHQGDGSLPIECLENDNNVQCNQOTGICSPISPYWSGIDC
SILPHTSAPSDPSGNSASIGWNNHTYLOVFIGNKQCIPIYINISKEIICAKP
PSTGLKSVLQNSINTKDIYOYNDKQCPNCHCTSTNGICNTTGTGVCIGRWS
GYDLSLGNPSSGDNSSGLPSTTIDGNTGNVINTNQNTQNTLITKLIEDPKN
QINELLKNSIISNLSKTYTITQIQNDKNTITIDEINKRDITTFAGINFKUT
SGVMKTVSIENYSSNLTQLOKMSVNEIQTNENDCNNDQIEINTPKQDLSL
NVTIKADKLVGRFIDRIEDSGITFMQTLVSKNSDITVGMPLPHTCKCLDPL
DFSVLVSPFVDECGDNDRKSWFLPVIVVPIVFLTLIIIFVITIKRFEVWYLPK
SKIKQLANLNK"

Join(11974..12125,12329..12522,12630..13054)
/note="ORF_ID:dd_02935"
/codon_start=1
/product="Putative GTPase"
/protein_id="AAM43760.1"
/db_xref="GI:21166143"
/translation="MDPLSISWESKVVLLGSSDVGTALSLRYVDGIPPKELTSTIGA
SFLTITINIEGNKIKLIWSDAGQDRFSLATLIYRGACVAILVDFDIQQKTFIVKGS
WYELKANIQBEIINVLCKIDLEQNRVZSEKLYADEINAMVYETSAKEHGE
GNFLIEGKLLILNKHNEYKFOQQQKLKHOQQOQSLYQYHIQQQKHQQQQQQQQ
QIFKQOQYNNGH5LQGSINHNQNSTYNSDSDCCG"
complement(join(13444..14192,14284..14623,14713..15015))
/note="ORF_ID:dd_02938"
/codon_start=1
/product="AMMONIUM TRANSPORTER AMTA"
/protein_id="AAM43761.1"
/db_xref="GI:21166144"
/translation="MVAGEIIKGVAEITNGSSSVQKYLDCANQVAPDFGNTTWWL
LSTILVGMPLAFPEAGLLRSKNTLSIITQIMSGIVLTVWMAQFYSLTGPDOK
GIIGNLDAFLINYSYDDCSNAPNIIPAAVAFMMMPANITPLMTGAPAEKVKFA
FIALTVAEIIIVFVVAHWINGGWLHKYFGVLDPAAGIVIHTSAGVSALVIALYVGR
RADFEKYGEFFPNLPLATIGALLMWMGFNAGSALAAGNATSAVASTQIGGSF
SAIVMILSAAKGPNTSVINGVAGITPASGINSQYSGIGICGLASYSYV
VLLKHLHIDALDVSSVHGLTGIIGSLAIGCAELSVNPNANGAFYGNPLGTOL
LCVSVAVMAAAWVLLKIIDATIGVKIDSEBELGLDVEHGEFAYHNISLQGNEN
HYSSVINSHDRFK"
15832..17823
/note="ORF_ID:dd_03054"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43762.1"
/db_xref="GI:21166145"
/translation="WKDILSPKYLHIDILKPIIDYLLTNNKINYNDDNNNNKVIK
NYLYINEIILKLSVSHWYQIISNLIISNLIENLEWPDLYSDVPEENYKGI
NFIAYDFNKIYDVTQSNFKQVQKSIKIKNGSVIEDIFRNQVLCIVLITVEKLL
QLSKLETGNLKYRKISKNPYSKDQKPLFGGIAKGIEKGLDIISSNNNNNN
NN
YDIYDQPIILYNTIDYEGSDDYNEKTFIRILYKLSKNIIPRSNTNENVHADFSLM
FNNKINNOVESIRIDSFVEPIDLQNVVELTNLHTLSIPLHFHOMILLQIGIKLHCR
GGDILLEKDNVQMDKMWIRSLLSAKIKNPKSNKCYTECINKSINSYNNENYSA
SRHVKLTFSNGIKQLSSSTFKFNSNMDIISIDTFSELKNNKTKLKYVKEKPHY
GIRLFTTSEEVAYQFNINLYLNIENFSGTSGSMLSNIRKHFSLAFPPYKRLI
FESLPKHYNYFQYSITVIYDITELLNEIINLSNLSKIVQOQQQQQQQQQIKFKI
VIERKFSNDRHFQSLANDRINGDGLFVITVSNKKLC"
complement(join(17912..18092,18308..18333))
/note="ORF_ID:dd_03055"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43763.1"
/db_xref="GI:21166146"
/translation="WTIIFLICLDASTQSTTNNNNNNNNNNNNNNNNNNNNNNNN
NN
18436..20349
/note="ORF_ID:dd_02950"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43764.1"
/db_xref="GI:21166147"
/translation="WKDILSPKYLHIDILKPIIDYLLTNNKINYNDDNNNNKVIK
NYLYINEIILKLSVSHWYQIISNLIISNLIENLEWPDLYSDVPEENYKGI
NFIAYDFNKIYDVTQSNFKQVQKSIKIKNGSVIEDIFRNQVLCIVLITVEK
LLQSLKLETGNLKYRKISKNPYSKDQKPLFGGIAKGIEKGLDIISSNNNNNN
NN
YNEKTFIRILEYKSNKIKRFTSVNHPADPSLFMNNKINNQVIESIDFVDPDQ
QNVLELTNLHTLSIPLHFHOMILLQIGTIELHCRGGDLEHTLNDMDQWDMQSLIA
SKSIKNPKIKNKYCLGECINQISIKYNNENYSAASROVKKLFSDGIKQLSSSTFK
KTYNMDIISIDTFSELKNNKTKLKYVKEKFLVNYCLRNKEVVAEHQFINLYLEN
VFSNGASGSMLSNNIKHFKFSLPDKLIPESLPKHYNYFQYSITVIYDITELLNE
ELLNEIINLSNLSKIVQOQQQQQQQSIKFKIIVINKINNNKHFQTLINDRINGDNG
LFVITVSNKKLC"
Join(122336..123207,23308..23806)
/note="ORF_ID:dd_02928"
/codon_start=1

/product="rac-family serine/threonine kinase homolog"
/protein_id="AA043765.1"
/db_xref="GI:21166148"
/translation="MTTISNNLPOLSEIPSTKLLITKQIKLRETSPPKSWMDTP
KNTIISPTPECESSSTITTPESLSASGHESSISDESKEITIIAILKLPNSVQ
DFTPLCNGAETIIVKISINSINLNFIIKELVKGFGKVFQVHVVDVTKVAKLVK
KNHIIAASVNTVALEKIDILKKSHPFIIVLHYAFQNEKCYLVMDVNGGOLFVHLO
KEAIFSEDQVRFVMAELIILALEHDSNIIVHRLDKPENILDSQGHILITDFGLAKLE
VKTNNITPSFAGTLYVMAPEMICHATGKAVDWSIGILMYDMWIKGKPFHKNRALM
OEKIISSEKAPKPFVSSARSILNGLLTKDPTKRLGANGAIEIKRHPFPKSIQWKIE
NKEITPPFVSTGIDDISNFDHASLKAQRDSFSTSLSSSQQAYPDGFSFVTPV
LLESQK"
complement(join(24089..24410,24517..24779))
/note="ORF ID:dd_02931"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA043766.1"
/db_xref="GI:21166149"
/translation="MSDPGEENVEITEFVEGDINENDLIDGNEYVDNGISPEITT
TFONSNNNNNNHNNNSYSGDGLSSVDGDMKPEAPAPVARYLEKHEKEMQEK
KXSEERQKKIABAKOSLONFYSEBEAKKKTALKNRDHNKSLSDTSGNTHTWES
VWSMIDLOAKPNPANKDTERMREILRLKNQPIV"
25279..26412
/note="ORF ID:dd_02934"
/codon_start=1
/product="Mitochondrial genome maintenance protein MGM101
precursor."
/protein_id="AA043767.1"
/db_xref="GI:21166150"
/translation="MKRFISTSLRLVSLNKNNNNNKFNKSIQVFNREFSSNEG
LNKVFNGNNSSSTPTKVLNDSQYQEDKQGISKEPFSKEIVDTLLADLNDDI
EIKDGLIYLPKLYRIINCAEPGQWALKPFPVVGKTLIRPVALYCLGRVAE
SIGEQVFNPSFISFATATSAKSNALVRCCKDLGSSLMDFPIRQWKSIAIYERW
CENSKTERLFLFNLNRSNQLPFWKESDFNFPSSNSISNNNSNNNN
NNNSDSSSINQOQETISYHODSSPSKITEQDDSEIDIDDDVPPQLKYAGK
TWQLVADPKGLVGLWASSSFPNPKIKSQATAILEFINKSQ"
complement(26482..27039)
/note="ORF ID:dd_03240"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA043768.1"
/db_xref="GI:21166151"
/translation="MFIYFEPFHYNNVTAIPDFKILTFVQVFNITIIIPCKEFLSKDTVT

Query Match 3.1%; Score 52.6; DB 2; Length 56956;
Best Local Similarity 46.6%; Pred. No. 0.54; Indels 0; Gaps 0;
Matches 169; Conservative 0; Mismatches 194

QY 1091 TTTTTCATATAAATAAGTAAATCTTTTTCCTAACCAATAAATAATATTGAAAATC 1150
| | | | |
DB 52626 TTTTTCATATAAATTTATTTTGTGTTTATTTTATTTTATTTTATTTTATTT 52685
| | | | |
QY 1151 TTTCCAACTAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
| | | | |
DB 52686 ATTATTTATTTTAAATTTATAAATAATAGAAAATATAAATAAATAAATAAACAAC 52745
| | | | |
QY 1211 ATTTTCATTGGCAGTGACTAGTAAGTAACTAACAGAACTAGTTCTGATTTT 1270
| | | | |
DB 52746 TTTTAAATTATAGTGGATTTTATTTTATTTTATTTTATTTATATAAAGGGGTGT 52805
| | | | |
QY 1271 GGATTTTGAAGCTTTCTTAGTTTAAAAAACAAGTATATTACTAAACAATAAAGAAAAA 1330
| | | | |
DB 52806 CCAAGTAAAGAGTTTATTAATGTAATGAAAATAAATAAATAAATAAATAAATAAATAA 52865
| | | | |
QY 1331 CATTTTGTGAAAGAGAAATAAAGTTTACTTGACCCCATTTGTACAGATGTCCTCAATA 1390
| | | | |
DB 52866 AATGTAAACAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 52925
| | | | |
QY 1391 ATACTGTAGAGATAGACGCAATGGAAGTCAATTTTGTTCAGTGGTACATCGCAATGGT 1450
| | | | |
DB 52926 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 52985
| | | | |
QY 1451 TCT 1453

Db 52986 TTT 52988

RESULT 13
AX461144/c
AX461144
LOCUS AX461144 2000 bp DNA linear PAT 08-JUL-2002
DEFINITION Sequence 73 from Patent WO0198480.
ACCESSION AX461144
VERSION AX461144.1 GI:21726352
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
AUTHORS Budworth, P., Brown, D., Chang, H.S., Zhu, T., Han, B., Wang, X. and
Cooper, B.
PROMOTERS for regulation of plant gene expression
TITLE Patent: WO 0198480-A 73 27-DEC-2001;
JOURNAL Syngenta Participations AG (CH)
FEATURES
Location/Qualifiers
1..2000
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 714 a 371 c 305 g 610 t
ORIGIN
Query Match 3.1%; Score 52.4; DB 6; Length 2000;
Best Local Similarity 47.1%; Pred. No. 0.98;
Matches 161; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 1075 AGTAGGGCTTACAGCTTTTCCATAAAATTAAGTAAATCTTTTTCCTAACCAATA 1134
| | | | |
DB 1019 ATTAGGTCGTATCTCTTTAACTATATAAATAAATAAATAAATAAATAAATAAATG 960
| | | | |
QY 1135 AAAATTATTGAATCTTCCACCATAGAAAAGTTAAATTTGATCAGGATGGAATTT 1194
| | | | |
DB 959 AGAAAATTGAAGTAGAAGTACGAAACAGGTCTTAAACACTGTTCTTATTTTATTTGATATCAT 900
| | | | |
QY 1195 TTGTACAAAGCTAGTATTTTTCATTTGGAGTGTACTAGTAACTAGTAAAGTAAACAGA 1254
| | | | |
DB 899 TTCAAACTTGAGATTTATTAGCAATTAACCTACACGAAAAAATAAATAAATAAATAAATG 840
| | | | |
QY 1255 ATGAGTTTCTGATTTTGGATTTTGAAGCTTTTCTAGCTTAAATAAATAAATAAATAAATAA 1314
| | | | |
DB 839 AAGTGTTTTTTTCTGTTTTTGTGTTTTTGAACACGTTAAAGAACCGAAACCTGATCAGTTATTA 780
| | | | |
QY 1315 AACATAAAGAAAAACATTTTGTGAAAAAGAAAAAATAAAGTTTACTGACCCCATTTGTAC 1374
| | | | |
DB 779 GCAAGCGAGCTACCTTTCTGTTTGTGACGATCCTATAATAAAGTTTTCACGGAAC 720
| | | | |
QY 1375 AGATGGTCCCAATAATAATACATGATAGAGATAGAGCAATGGA 1416
| | | | |
DB 719 GCCTCTGACGGTTAATCGCATCGGTAGAGAAAGATCGGTGGA 678
| | | | |
RESULT 14
AC004401/c
AC004401
LOCUS AC004401.3 GI:20197125
DEFINITION Arabidopsis thaliana chromosome 2 clone F21P24 map C1C06C07,
complete sequence.
ACCESSION AC004401
VERSION AC004401.3
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 94038)
REFERENCE 1 (bases 1 to 94038)

Db	7022	TATATTAATAAAAAATATATTATTATATAAAATATATTATTATTTTAAATTTTGTGATTTTCT	7081
Qy	1274	TTTTGAAGCTTTTCTTAGGTTAAABAAACAGCTATATTACTATAACATATAAAGAAACAT	1333
Db	7082	TTTATTTCTTTAAATTAATATATTATAAATAATTTATTTACTATAATAAATTTATATCAT	7141
Qy	1334	TTTGTGAAAGAGAAATAAAGTTTACTGGACCCCAATTGTACAGATGGTCCCATATAATA	1393
Db	7142	AAAATTAATAAAAAAGTAATACTATATAATAAATAAATAATTTATCTAAATCAACAAAAAACCT	7201
Qy	1394	CTGATAGAGATAGA	1408
Db	7202	TGGAACCAAAAAAGGA	7216

Search completed: February 2, 2004, 00:53:01
Job time : 6372.59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

un on: February 1, 2004, 21:36:05 ; Search time 491.261 Seconds
(without alignments)
9214.976 Million cell updates/sec

file: US-09-938-842A-3729

Perfect score: 1677
Sequence: 1 ggttaaggggttttctatg.....ttttctcgcgtatatttta 1677

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_19Jun03.*

```
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1677	100.0	1677	24	Arabidopsis thalia
2	242.4	14.5	1483	21	Arabidopsis thalia
3	50.6	3.0	8056	22	Haematopoietic cel
4	49.6	3.0	422	25	Human breast cance
5	49	2.9	369	22	Human cervical can
6	48.6	2.9	20933	24	Human angiogenesis
7	47.8	2.9	2958	24	Arabidopsis thalia
8	46.6	2.8	516	25	Bovine EST associa

9	46.4	2.8	7047	24	ABK28385	DNA transcription
10	46.4	2.8	17848	22	AA545323	Chemically pretrea
11	46.4	2.8	17848	24	ABK39976	Human chemically p
12	46.4	2.8	17848	24	ABK28164	DNA transcription
13	46	2.7	8056	25	ABZ10246	Haematopoietic cel
14	45.6	2.7	37515	24	ABQ66998	Human angiogenesis
15	45.4	2.7	6352	24	ABL70563	Chemically treated
16	45.4	2.7	6352	24	AA561235	Human gene regulat
17	45.4	2.7	6352	24	ABK31340	Signal transductio
18	44.8	2.7	1271	24	ABZ15024	Arabidopsis thalia
19	44.6	2.7	2000	24	ABZ16233	Arabidopsis thalia
20	44.2	2.6	7749	24	ABL70435	Chemically treated
21	44.2	2.6	8711	22	AA545700	Tumour suppressor
22	44.2	2.6	19734	24	ABL33933	Human immune syste
23	44.2	2.6	113515	24	ABL34175	Human immune syste
24	44	2.6	468	23	ABV07439	Human prostate exp
25	44	2.6	6636	24	ABN80114	Human chemically m
26	44	2.6	7498	24	ABL32257	Human immune syste
27	43.6	2.6	10710	24	ABN79996	Human chemically m
28	43.6	2.6	11473	24	ABK40030	Human chemically p
29	43.6	2.6	11473	24	ABL33355	Human immune syste
30	43.6	2.6	14752	20	AA20256	Borrelia burgdorfe
31	43.4	2.6	6533	24	ABK28383	DNA transcription
32	43.4	2.6	7008	22	AA545532	Tumour suppressor
33	43.4	2.6	15161	24	ABL70458	Chemically treated
34	43.4	2.6	15161	24	AA561423	Human gene regulat
35	43.4	2.6	64081	24	ABA92787	Buchnera sp. genom
36	43.2	2.6	6175	24	ABL33307	Human immune syste
37	43	2.6	6244	24	ABL34485	Human immune syste
38	43	2.6	7244	22	AA546726	Tumour suppressor
39	43	2.6	40324	24	ABQ67150	Human angiogenesis
40	42.8	2.6	5682	24	ABL32572	Human immune syste
41	42.8	2.6	5682	24	ABL34500	Human metastasis a
42	42.8	2.6	7809	25	ABZ03977	Haematopoietic cel
43	42.8	2.6	7809	25	ABZ10123	Haematopoietic cel
44	42.8	2.6	8056	25	ABZ10100	Haematopoietic cel
45	42.8	2.6	8056	25	ABZ10100	Haematopoietic cel

ALIGNMENTS

RESULT 1
ABZ15924
ID ABL15924 standard; DNA; 1677 BP.
XX ABZ15924;
AC ABZ15924;
XX 21-JAN-2003 (first entry)
DT
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 3729.
XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX
XX WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
XX 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX (SCRI) SCRIPPS RES INST.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Harper JF, Kreps J, Wang X, Zhu T;
XX WPI; 2002-304127/34.

19-FEB-2001 (first entry)
 Arabidopsis thaliana SBE1 cDNA.
 SBE1; shrunken seed gene; storage reserve; storage protein;
 oil body; transgenic plant; ss.
 Arabidopsis thaliana.
 Key Location/Qualifiers
 CDS 122..1225
 /*tag= a
 WO200061735-A1.
 19-OCT-2000.
 07-APR-2000; 2000WO-US09192.
 08-APR-1999; 99US-0128651.
 (GEO) GEN HOSPITAL CORP.
 Lin Y;
 WPI; 2000-679483/66.
 P-PSDB; ABL9718.
 Novel shrunken seed gene useful for producing transgenic plants having
 altered production of food storage reserve material, intracellular
 transport of storage protein and formation of protein or oil bodies -
 Claim 8; Page 57; 64pp; English.
 The present sequence is that of Arabidopsis thaliana SBE1 (shrunken
 seed) cDNA, which was isolated from a seedling cDNA library using
 a partial genomic clone as probe. SBE1 encodes a protein (see
 AAB19718) that, when expressed in a cell of a plant, modifies or
 alters the production of a food storage reserve material (e.g.
 protein, lipid or carbohydrate storage reserve), facilitates the
 intracellular transport of a storage protein, or facilitates the
 formation of protein or oil bodies. The invention provides a
 transgenic plant (or plant cell, plant tissue, plant organ or
 plant component) which includes a recombinant SBE1 transgene that
 modifies the production of food storage reserves, thereby
 increasing nutritional value. An antisense construct is useful for
 modifying desiccation tolerance.
 Sequence 1483 BP; 461 A; 268 C; 347 G; 407 T; 0 other;
 Query Match 14.5%; Score 242.4; DB 21; Length 1483;
 Best Local Similarity 99.6%; Pred. No. 3e-46;
 Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 2Y 1 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATTCATTTGTTGAATGC 60
 2b 1226 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATTCATTTGTTGAATGC 1285
 2Y 61 TTUUTTCAGATCATCAAGGCTCTACAGATTTCTTGGGAAGTTTTCAGGCTTTGTTA 120
 2b 1286 TTUUTTCAGATCATCAAGGCTCTCTACAGATTTCTTGGGAAGTTTTCAGGCTTTGTTA 1345
 2Y 121 GAAATTTGTTTATTGCAACAGGTAGAGACATCAACATAGACAGATGTATCTGAAGAGA 180
 2b 1346 GAAATTTGTTTATTGCAACAGGTAGAGACATCAACATAGACAGATGTATCTGAAGAGA 1405
 2Y 181 TAGCTTCTTATGTTTAAAGAAATGCGGATACCAATAAACAGCATCAATTAAGAT 240
 2b 1406 TAGCTTCTTATGTTTAAAGAAATGCGGATACCAATAAACAGCATCAATTAAGAT 1465
 2Y 241 TAAA 244
 2b 1466 TAAA 1469

RESULT 3
 ABZ10246
 ID ABZ10246 standard; DNA; 8056 BP.
 AC ABZ10246;
 XX
 XX 16-JAN-2003 (first entry)
 DT
 XX Haematopoietic cell proliferation disorder related DNA sequence #386.
 DE
 XX Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200277272-A2.
 PN
 XX 03-OCT-2002.
 PD
 XX 26-MAR-2002; 2002WO-EP03401.
 PF
 XX 26-MAR-2001; 2001US-278333P.
 PR
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
 PI Pelet C, Schwöpe I, Ziebarth H;
 XX
 XX WPI; 2003-018942/01.
 DR
 XX Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent
 PT that distinguishes between methylated and non-methylated CpG
 PT dinucleotides -
 XX
 XX Claim 28; SEQ ID 386; 117pp; English.
 PS
 XX The present invention describes a method for detecting and
 CC differentiating between hematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used for
 CC differentiating between healthy hematopoietic cells and proliferative
 CC disorder hematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of hematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of hematopoietic cell proliferation disorder related
 CC DNA sequences. The nucleotide sequences from the present invention can
 CC also be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC hematopoietic cell proliferative disorders. The present method enables
 CC a highly specific classification of hematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients.
 XX
 XX Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;
 SQ
 Query Match 3.0%; Score 50.6; DB 25; Length 8056;
 Best Local Similarity 49.4%; Pred. NO. 0.088;
 Matches 131; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
 1091 TTTTCCATAAAATTAAGTAATCTTTTTCCTTACCAATATAAAATTTGAAATC 1150

Db 5627 TTTTGTGTTAAATTTTAAATAATTTTATTTTTCATATAAAAAAATAAATAATG 5686
Qy 1151 TTCCCAACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGT 1210
Db 5687 ATTGAATTTGATGATATGATGTTGTAATATATATTTGAAATTTTAAATGTTAAA 5746
2y 1211 ATTTCAATTTGGGAGTGACTAGTAAGTACTAACCAGAAAGTCTTCTGATTTT 1270
Db 5747 TTTTGTGTTATATAAAAAATGATAAATGAAATTAATGAAATAAATTTTATTTT 5806
Qy 1271 GGATTTTGAAGCTTTCTTGGTTAAATAACAAGTATATTACTAAACAATAAAGAAAA 1330
Db 5807 TTTTATGTTTATTTTAAATAATTTGAAAAAATGTTATATAAAAAAATAAATA 5866
Qy 1331 CATTTGTGAAAAGAGAAATPAAAGT 1355
Db 5867 AAAAAATAAAAAAATAATTAATT 5891

RESULT 4

AAL15924/c
ID AAL15924 standard; cDNA; 422 BP.

XX AC AAL15924;

XX DT 07-DEC-2001 (first entry)

XX DE Human breast cancer expressed polynucleotide 8381.

XX KW Human; breast cancer; cell marker; cytostatic; ss.

XX OS Homo sapiens.

XX PN W0200151628-A2.

XX PD 19-JUL-2001.

XX PF 10-JAN-2001; 2001WO-US00798.

XX PR 14-JAN-2000; 2000US-0176077.

XX PR 14-MAR-2000; 2000US-0189167.

XX PR 24-MAR-2000; 2000US-0192099.

XX PR 29-MAR-2000; 2000US-0193480.

XX PR 15-MAY-2000; 2000US-0205230.

XX PR 09-JUN-2000; 2000US-0211315.

XX PR 25-JUL-2000; 2000US-0220534.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX DR WPI; 2001-451856/48.

XX PT New peptide useful as a marker for the diagnosis of breast cancer -

XX PS Claim 1; Page 1516; 3695pp; English.

XX CC The invention relates to human breast cancer expressed polynucleotides (AAL07544-AAL26789) and methods of assessing whether a patient is afflicted with breast cancer by examining the correlation between the expression of certain markers and the cancerous state of breast cells. The polynucleotides and encoded polypeptides are potential markers for detecting, diagnosing, monitoring, characterizing treating and potentially preventing breast cancer. The polynucleotides and encoded polypeptides are also useful for isolating compounds with cytostatic activity.

XX SQ Sequence 422 BP; 112 A; 32 C; 25 G; 189 T; 64 other;

Query Match 3.0%; Score 49.6; DB 22; Length 422;

Best Local Similarity 42.1%; Pred. No. 0.072;

Matches 118; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Qy 1088 AGCTTTTCCATAAAAATAAGTAAATCTTTTTTGCCTAACCAATAAAAAATTATTGAAA 1147
Db 331 ATNNNTTTGAAAAAATAAATAAATAAATTTGNGGAAATTTAAAAAATTTTNTTAAAAA 272
Qy 1148 ATCTTTCCCAACCATAGAAAAGTTAAATTTTGATCAGCGATGGAATTTTGTACAAAGCTA 1207
Db 271 GNNNNCANAATTAATANTTTTNGTTTNAAAAAAANAANANTTTTATATTTT 212
Qy 1208 GGATTTTCATTTGGGAGTGACTAGTAAGTACTAACCAGAAAGTCTGAT 1267
Db 211 AAGNTTTTANNANANATTTNNNANNTTAAATTTTNTTAAANANTAAANNCNTTTTT 152
Qy 1268 TTTGGAATTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTACTAAACAATAAAGAA 1327
Db 151 TTTTNTTTTATTAANNNNNNAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 92
Qy 1328 AACATTTTGTGAAAAGAGAAATAAAGTTTACTGTGACCCC 1367
Db 91 AAAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAAGTACCTC 52

RESULT 5

AAH69544/c

ID AAH69544 standard; cDNA; 369 BP.

XX AC AAH69544;

XX DT 19-SEP-2001 (first entry)

XX DE Human cervical cancer marker nucleic acid 818.

XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX PN W0200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33312.

XX PR 08-DEC-1999; 99US-0159681.

XX PR 21-DEC-1999; 99US-0171350.

XX PR 14-MAR-2000; 2000US-0189315.

XX PR 12-MAY-2000; 2000US-0203791.

XX PR 09-JUN-2000; 2000US-0210600.

XX PR 21-JUL-2000; 2000US-0220114.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX DR WPI; 2001-375006/39.

XX PT New isolated nucleic acid for diagnosing and treating cervical cancer and for assessing and detecting compounds for treating the cancer -

XX PS Claim 1; Page 242; 1051pp; English.

XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with cervical cancer with cytostatic activity. The nucleic acids and encoded polypeptides are useful to assess if a patient is afflicted with cervical cancer or has a pre-malignant condition; to monitor the progression of cervical cancer or a premalignant condition in a patient; and to select and/or assess the efficacy of a compound or therapy for inhibiting cervical cancer in a patient. The nucleic acids may also be useful for gene therapy.

XX SQ Sequence 369 BP; 96 A; 28 C; 29 G; 172 T; 44 other;

Query Match 2.9%; Score 49; DB 22; Length 369;

Best Local Similarity 44.4%; Pred. No. 0.096;

Matches 115; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

Y 1091 TTTTCCATAAAATTAAGTAAATCTTTTGGCTAACCAATAAAAATTATTGAAATC 1150
b 313 TTGTTAAAAAACAACCAAAAGAGTNTTTTTCGGTAGGAAAAAATTTTANTTG 254
Y 1151 TTTCCACCATAGAAAGTTAAATTTGATCAGCATGGAATTTTGTACAAAGCTAGGT 1210
b 253 ATTCATCCNNGCATTTGTTTTTANGTTTTTNCAGCGGNNNTTTGGANAAAAAGTNA 194
Y 1211 ATTTTCATTTGGAGTGTACTAGTAACTAGTAACTAACACAGATGAGTTTCTGATTTT 1270
b 193 NTTTTTCNTTTGTTTTTAATAAANNAANNAATTTNAAAAAANNNNNNTT 134
Y 1271 GGAATTTGAGCTTTCTTAGTTTAAAAAACAAGTATATTTACTAAACAATAAAAAGAAAA 1330
b 133 NAAAAANTTAATTTTTTTTNNAAAAAAGTCCAAAAATNNAAANNGAAAA 74
Y 1331 CATTTTGTGAAAAAGAAA 1349
b 73 AAAAAAANAAAAA 55

RESULT 6
BQ67124
D BQ67124 standard; DNA; 20933 BP.
X X
C BQ67124;
X X
T 28-AUG-2002 (first entry)
X X
E Human angiogenesis associated polynucleotide SEQ ID NO 154.
X X
W Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
W inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
W macular degeneration; inflammatory bowel disease; Crohn's disease;
W antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
W antiarteriosclerotic; ds.
S Homo sapiens.
X X
N WO200246454-A2.
X X
D 13-JUN-2002.
X X
F 06-DEC-2001; 2001WO-EP14320.
X X
R 06-DEC-2000; 2000DE-1061338.
X X
A (EPIC-) EPIGENOMICS AG.
X X
I Schacht O;
X X
R WPI; 2002-500450/53.
X X
T New nucleic acid fragments from chemically treated
T angiogenesis-associated genes, useful for determining methylation
T status, e.g. in diagnosis or treatment of cancer -
S Claim 1; SEQ ID NO 154; 41pp + Sequence Listing; German.
X X
C The invention relates to a nucleic acid (I) comprising a segment of 19
C bases of chemically pretreated DNA of angiogenesis-associated genes (II)
C having sequences (ABQ6971-ABQ67178) or their complements. (I), also
C related oligomers, are used to evaluate the methylation status and/or
C single-nucleotide polymorphisms, in angiogenesis-related genes, for
C diagnosis and treatment of eye diseases, proliferative retinopathy,
C neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
C diabetic retinopathy, macular degeneration caused by neovascularisation,
C psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
C Crohn's disease.
C Note: The sequence data for this patent did not form part of the printed
C specification, but was obtained in electronic format directly from WIPO
C at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 20933 BP; 5401 A; 396 C; 5174 G; 9962 T; 0 other;
Query Match 2.9%; Score 48.6; DB 24; Length 20933;
Best Local Similarity 56.6%; Pred. No. 0.32;
Matches 90; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 1091 TTTTCCATAAAATTAAGTAAATCTTTTGGCTAACCAATAAAAATTATTGAAATC 1150
Db 1180 TTTTATATAATTAAGATATAATTTTTTTAGAAAAATAGGTAAAGATATATAATGG 1239
QY 1151 TTTCCACCATAGAAAGTTAAATTTGATCAGCATGGAATTTTGTACAAAGCTAGGT 1210
Db 1240 TTTATPAAAGATATATTTTTTTATTAGATGGAATAGTTATTTTAAATATTGTTAGTT 1299
QY 1211 ATTTTCATTTGGAGTGTACTAGTAACTAGTAACTAGTAACTAGTAA 1249
Db 1300 TTTTAAGTTGGAGTATAATAGTAAATTTTAAATTTAAA 1338
RESULT 7
ABZ12533/c
ID ABZ12533 standard; DNA; 2958 BP.
XX
AC ABZ12533;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 338.
DE Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
KW Arabidopsis thaliana.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
PT
XX Claim 144; SEQ ID NO 338; 577pp + Sequence Listing; English.
PS The invention relates to identifying a stress condition to which a plant
PS cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 2958 BP; 828 A; 621 C; 576 G; 933 T; 0 other;

Query Match	2.9%;	Score 47.8;	DB 24;	Length 2958;
Best Local Similarity	50.7%;	Pred. No. 0.3;		
Matches 115;	Conservative 0;	Mismatches 112;	Indels 0;	Gaps 0;
QY	1228	ACTAGTAAC	TAGTAAGTACTAAC	CAGAATCAGTTC
				TCTGATTTTGGATTTTGAAGCTTTTC
Db	410	ATTAGCACT	AGCTAGTAATTAAGTGT	TTTAGTGTACAAAGTATTTTACAGGGTTC
				TATC 351
QY	1288	TTAGGTTAAAAA	CACAGTATATTACT	TAAACAATAAAGAAAAACATTTTGTGAAAAAGAGA
				1347
Db	350	ATAAAATAAAA	GAGGATTAATATTACT	AAATAGCCAAAAAACACACGTAAAAATAGCAGA
				291
QY	1348	AAATAAAGTTT	TACTGGAGCCCATTTGT	CAGATGGTCCCATATAATCTGATAGAAGATAG
				1407
Db	290	GAGAAAGGTTT	TGTGCAANTTAGTCAT	GAACTCACTTATCAACAGGCTGAATGAC
				231
QY	1408	AGCAATGGAAG	TGATTTGTTCTACGGTGT	CAATCGGAATGGTTCTT
				1454
Db	230	AAAACAAATCC	TGTTGTTGCAAGTTT	CAACAGTAGCGAGACACTT
				184

RESULT 8	
ABX40620/c	
ID	ABX40620 standard; cDNA; 516 BP.
XX	
AC	ABX40620;
XX	
DT	20-FEB-2003 (first entry)
XX	
DE	Bovine EST associated with lactation/muscle/fat deposition #5785.
XX	
KW	Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
KW	muscle deposition; fat deposition; genome mapping; gene identification;
KW	gene analysis; cattle breeding.

XX	US	BOS laurus.
XX	PN	US2002137139-A1.
XX	XX	
XX	PD	26-SEP-2002.
XX	XX	
XX	PF	24-SEP-2001; 2001US-0960352.
XX	XX	
XX	PR	12-JAN-1999; 99US-115707P.
XX	PR	11-JAN-2000; 2000US-0480902.
XX	XX	
XX	PA	(BYAT/) BYATT J C.
XX	PA	(MATH/) MATHIALAGAN N.
XX	PA	(TACN/) TAO N.
XX	PA	(WARR/) WARREN W C.
XX	XX	
XX	PI	Byatt JC, Mathialagan N, Tao N, Warren WC;
XX	PI	WPI; 2003-110599/10.
XX	DR	
XX	XX	
XX	PT	New nucleic acid associated with lactation, and muscle and fat
XX	PT	deposition, useful for genome mapping, gene identification and
XX	PT	analysis, cattle breeding, or for genetically improving cattle
XX	XX	
XX	PS	Claim 2; SEQ ID No 5785; 245pp; English.

The invention relates to a purified nucleic acid molecule associated with lactation or muscle and fat deposition (designated LMFD), derived from cattle, and the LMFD nucleic acid can specifically hybridize to a second nucleic acid molecule comprising any of 1512 nucleotide sequences, appearing as ABX3436-ABX49947, or complements of them. Also included are: (1) a transformed cell having a nucleic acid comprising an LMFD nucleic acid linked to a promoter and a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and (2) determining a level or pattern of a molecule in a bovine cell or tissue comprising: (a) incubating a marker

CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridization between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/cgi-bin/sequence.html?DocID=200201371139.

```

XX
SQ      Sequence 516 BP; 79 A; 14 C; 18 G; 403 T; 2 other;
      Query Match      2.8%; Score 46.6; DB 25; Length 516;
      Best Local Similarity 46.2%; Pred. No. 0.37;
      Matches 154; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

Qy      1098 ATAAATTAAGTAATCTTTTTTTCGCTAACCAATATAAAATATTATTGAAAAATCTTTCCCAA 1157
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      434 ATAAAAAATAAAAAATAAAAAAATAAAAAATAAAAAATAAAAAAATAAAAAAATAAAAAA 375

Qy      1158 CCATAGAAAAGTTTAAATTTCGATCAGCGATGGAAATTTTGTGTACAAAGCTAGGTATTTCAT 1217
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      374 TAATAAAAAAATAAAAAATAAATAAAAAACAATAATTTATAAAAAAATAAAAAATAAAAAA 315

Qy      1218 TTGGGAGTCTACTAGTAACTAGTAGTACTAAACGAGATCGAGTTCTGTATTTGGATTTT 1277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      314 AAAAAAATAAAATAAAAAATAAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAA 255

Qy      1278 GAAGCTTTTCTAGGTTAAAAACAAGTATATTACTAAACAATAAAGAAAAACATTTTG 1337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      254 AAAAAAATAATAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAATAA 195

Qy      1338 TGA AAAAGAGAAATAAAGTTTCTAGTGGACCCCATTTGTACAGATGGTCCCATATAATACTGA 1397
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      194 AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAATAA 135

Qy      1398 TAGAGATAGACAAATGGAAAGTGATTTGTTCA 1430
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      134 AAAAAATAAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAA 102

```

RESULT 9	
ABK28385	ABK28385
ID	ABK28385 standard; DNA; 7047 BP.
XX	
AC	ABK28385;
XX	
DT	23-APR-2002 (first entry)
XX	
DE	DNA transcription associated genomic DNA #130.
XX	
KW	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW	PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW	single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW	viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW	immunological disorder; Werner syndrome; developmental disorder;
KW	psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW	neurodegenerative disorder; Wardenburg syndrome; Niemann-Pick disease;
KW	myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW	angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW	polyglutamine disorder; solid tumour.
XX	
OS	Unidentified.
XX	
PN	WO200192565-A2.
XX	

06-DEC-2001.

06-APR-2001; 2001WO-EP03973.

06-APR-2000; 2000DE-1019058.

07-APR-2000; 2000DE-1019173.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;
WPI; 2002-090045/12.

New nucleic acids or oligomers, useful for diagnosing or treating diseases associated with DNA transcription, e.g. immunological disorders, Werner syndrome, psoriasis, myocardial infarction, solid tumours or cancer -

Claim 1; SEQ ID No 259; 32pp; English.

The invention relates to a nucleic acid, which comprises a segment of the chemically pretreated DNA of genes associated with DNA transcription from one of 346 sequences, and an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical to the chemically pretreated DNA of genes associated with DNA transcription. The set of oligomer probes are useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA. The nucleic acids are useful for diagnosing or treating diseases associated with DNA transcription (particularly with the methylation status), e.g. adenosine deaminase deficiency, viral infection, retroviral infection, Seary syndrome, haematological disorders, immunological disorders, Werner syndrome, tuberculosis, developmental disorders, psoriasis, Rieger's syndrome, neurological disorders, neurodegenerative disorders, Waardenburg syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial infarction, hypertension, angiogenesis, erythropoiesis, congenital heart disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours or cancer. Sequences ABK28127-ABK28472 represent DNA transcription associated genomic DNA molecules of the invention.

Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from the European Patent Office.

Sequence 7047 BP; 2247 A; 75 C; 1573 G; 3151 T; 1 other:

Query Match	2.8%	Score 46.4	DB 24	Length 7047
Best Local Similarity	49.7%	Pred. No. 0.8		
Matches 145	Conservative 0	Mismatches 146	Indels 1	Gaps 1
1084	TTACAGCTTTTCCATAAAATAAAGTAATACCTTTTTTGGCTCAACCAATAAAAAATTATT	1143		
3514	TTATATATGATATATGTTATGTTTAGGAATTTGTAATTTTGGATATTATTTTAGGAAT	3573		
1144	GAAAACTTTCCAAACCATAGAAAGTTAAATTTGATCAGCGAGTGGAAATTTTGTGTACAAA	1203		
3574	GAAAAATTCGGTTTATATA-AAAAATTTGATACGAATGTTTATAGTAGTATTTATTGTGAA	3632		
1204	GCTAGGTATTTTCATTTGGGAGTGTACTAGTACTAGTAAGTACTAACACAGAAATCAGTTTC	1363		
3633	TGGTCGAAATTTCTAGGTAAATTTATATATTTTTTTTTTAGGTGTTTAAAAATAAATTTGGT	3692		
1264	TGAATTTTGGATTTTGAAGCTTTCTTAGGTGTTAAAAAACAAGTATATTACTAAACAATAAA	1323		
3693	TTATGATATTAATAGGAATTTTGTATTTTGTATAAAATTTGAATTTATGATATTATATA	3752		
1324	AGAAAAACATTTTGTGAAAAGAGAAATAAAAGTTTACTGGACCCCATTTGTACA	1375		
3753	ATAATTTTAAATAAATTTTGAAGGAATTAATGTTTGAGTGAAATAATTTATAAA	3804		

RESULT 10

AA845323/c	
ID	AA845323 standard; DNA; 17848 BP.
XX	
AC	AA845323;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Chemically pretreated complementary DNA associated with cell cycle #14.
XX	
KW	Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW	human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW	grat-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW	arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW	immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW	PCR primer.
XX	
OS	Homo sapiens.
XX	
PN	WO200168911-A2.
XX	
PD	20-SEP-2001.
XX	
PF	15-MAR-2001; 2001WO-EP02945.
XX	
PR	15-MAR-2000; 2000DE-1013847.
PR	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019173.
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPiG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
DR	WPI; 2001-602751/68.
XX	
PT	Designing primers and probes for analysing diseases associated with
PT	cytosine methylation state e.g. arthritis, cancer, aging,
PT	arteriosclerosis comprising fragments of chemically modified genes
PT	associated with cell cycle -
XX	
PS	Claim 1; SEQ ID No 28; 28pp; English.
XX	
CC	Sequences AA845296-AA845520 represent chemically pretreated genomic DNA
CC	molecules associated with the cell cycle and specific PCR primers of the
CC	invention. The sequences are useful for detecting the methylation state
CC	of all CpG dinucleotides in a sequence and therefore for analysing
CC	associated diseases. By analysing cytosine methylations in the pretreated
CC	DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC	of existing diseases or the predisposition to specific diseases can be
CC	ascertained. The parameters may be compared to another set of genetic
CC	and/or epigenetic parameters, the differences serving as basis for
CC	diagnosis and/or prognosis events which are disadvantageous to patients.
CC	The sequences of the invention are useful for the diagnosis and therapy
CC	of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC	aging, glomerular disease, Lewy body disease, arthritis,
CC	arteriosclerosis, solid tumours and cancers.
XX	
SQ	Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 6 other;
XX	
Query Match	2.88; Score 46.4; DB 22; Length 17848;
Best Local Similarity	45.68; Pred. No. 1;
Matches 164; Conservative	0; Mismatches 196; Indels 0; Gaps 0;
QY	1088 AGCTTTTCCATAAATTAAGTAAATCTTTTTTTCGCTAACCAATTAATAATTATGAAA 1147
DB	10047 AACTTAAAAAAAACAATTTTAAAAAATTTAAAAAATTAATAAATTAATAAATTCAA 9988
QY	1148 ATCTTTCCACCTAGTAAAGTTAAATTTGATCAGGATGGAATTTTGTACAAAGCTA 1207
DB	9987 AAAAAATAAAAAATAAAAAAATAAATAAATACTAAAAATTTTAAAAAATAAATTC 9928
QY	1208 GGTATTTTCATTTGGGAGTGTACTAGTAACTAGTAACTAGTAACTAACAGATGAGTTCTGAT 1267

PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-090046/12.
XX New nucleic acids or oligomers, useful for diagnosing or treating
XX diseases associated with DNA transcription, e.g. immunological
XX disorders, Werner syndrome, psoriasis, myocardial infarction, solid
XX tumours or cancer
XX
XX Claim 1; SEQ ID No 38; 32pp; English.
XX
XX The invention relates to a nucleic acid, which comprises a segment of the
XX chemically pretreated DNA of genes associated with DNA transcription from
XX one of 345 sequences, and an oligomer, in particular an oligonucleotide
XX or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
XX to the chemically pretreated DNA of genes associated with DNA
XX transcription. The set of oligomer probes are useful for detecting the
XX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
XX in a chemically pretreated genomic DNA. The nucleic acids are useful for
XX diagnosing or treating diseases associated with DNA transcription
XX (particularly with the methylation status), e.g. adenovirus deaminase
XX deficiency, viral infection, retroviral infection, Sezary syndrome,
XX haematological disorders, immunological disorders, Werner syndrome,
XX tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
XX neurological disorders, neurodegenerative disorders, Waardenburg
XX syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
XX infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
XX disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
XX or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
XX associated genomic DNA molecules of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 6 other;
XX
XX Query Match 2.8%; Score 46.4; DB 24; Length 17848;
XX Best Local Similarity 45.6%; Pred. No. 1;
XX Matches 164; Conservative 0; Mismatches 196; Indels 0; Gaps 0;
XX
XX 1088 AGCTTTTCATAAATTAAGTAATTCCTTTTTCCTTAAACATAAAATATTGAAA 1147
XX 10047 AACTTAAATAAAACATATTTTAAATATTTAAATAATTTAAATTTTAAATTTCAA 9988
XX
XX 1148 ATCTTTCCACCATGAAAGTTTATTTGATCAGCGTGGAAATTTTGTACAAAGCTA 1207
XX 9987 AAAAAATAAAATAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATC 9928
XX
XX 1208 GGTATTTCATTTGGGAGTGTACTAGTAACTAGTAACTAACCCAGATGAGTTTCTGAT 1267
XX 9927 ATAAATAATACTAAATTAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATC 9868
XX
XX 1268 TTGGATTGTGAGCTTTCTTAGGTAAATAAAACAGTATATTTACTAACATAAAAGAA 1327
XX 9867 TATAAAAAAATAAAACATATTTTAAATTAACAAAAATAATCTAATAAAAAAATAAA 9808
XX
XX 1328 AAACATTTTGTGAAAGAGAAATAAAGTTTACTGGACCCCATTTGACAGATGGTCCCAT 1387
XX 9807 AATTAATAATCGAACAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATA 9748
XX
XX 1388 ATAACTACTGATGAGCAATGAGCAATGGAAGTATTTGTTCCAGTGGTACATCGGAAT 1447
XX 9747 AAATCTATTACCCATAAAAAATAAATTTAATCTTAACTAAAAACCAATTCATCCCTAAT 9688
XX
XX RESULT 13
XX ABZ10246/c
XX ID ABZ10246 standard; DNA; 8056 BP.
XX AC
XX ABZ10246;
XX
XX 16-JAN-2003 (first entry)
XX

DE Haematopoietic cell proliferation disorder related DNA sequence #386.
XX Human; haematopoietic cell proliferation disorder; cytostatic;
XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
XX cytosine methylation state; gens; ds.
XX
XX Homo sapiens.
XX
XX WO2002727272-A2.
XX
XX 03-OCT-2002.
XX
XX 26-MAR-2002; 2002WO-EP03401.
XX
XX 26-MAR-2001; 2001US-278333P.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
XX Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
XX Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
XX Pellet C, Schwöpe I, Ziebarth H;
XX
XX WPI; 2003-018942/01.
XX
XX Detecting and differentiating between hematopoietic cell proliferative
XX disorders, comprises contacting a target nucleic acid with a reagent
XX that distinguishes between methylated and non-methylated CpG
XX dinucleotides -
XX
XX Claim 28; SEQ ID 386; 117pp; English.
XX
XX The present invention describes a method for detecting and
XX differentiating between hematopoietic cell proliferative disorders
XX associated with at least 1 gene and/or their regulatory regions in a
XX subject. The method comprises contacting a target nucleic acid in a
XX biological sample obtained from the subject with at least 1 reagent,
XX which distinguishes between methylated and non-methylated CpG
XX dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
XX represent specifically claimed nucleotide sequences from the present
XX invention. Oligonucleotides from the present invention can be used: for
XX differentiating between healthy hematopoietic cells and proliferative
XX disorder hematopoietic cells; for differentiating between acute
XX lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
XX determining the cytosine methylation state and/or single nucleotide
XX polymorphisms (SNPs) of hematopoietic cell proliferation disorder
XX related sequences and their complements; and as primers for the
XX amplification of hematopoietic cell proliferation disorder related
XX DNA sequences. The nucleotide sequences from the present invention can
XX also be used for detecting a predisposition to, differentiation between
XX subclasses, diagnosis, prognosis, treatment and/or monitoring of
XX hematopoietic cell proliferative disorders. The present method enables
XX a highly specific classification of hematopoietic cell proliferative
XX disorders allowing for improved and informed treatment of patients.
XX
XX Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;
XX
XX Query Match 2.7%; Score 46; DB 25; Length 8056;
XX Best Local Similarity 50.5%; Pred. No. 1;
XX Matches 112; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
XX
XX 1137 AATTATTGAAAACTTTTCCACCATGAGAAAGTAAATTTGATCAGCGATGGAATTTT 1196
XX 5212 AAATATTAATTTTTCAAAAATAAAATTTAAATAATATCAAAATTTTATTTT 5153
XX
XX 1197 GTACAAAGCTAGGTATTTCAATTTGGGAGTGTACTAGTAACTAGTAACTAGCAAT 1256
XX 5152 TTAATTTATTTTATTTATTTTAAATAATATCAATATTTTAAATAATTT 5093
XX
XX 1257 GAGTTTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACCAAGTATATTACTAAA 1316
XX 5092 TTTTATTTTATTTTACATTTTAAATTTTTCCTTTTCAAAATAAAATTTTATTTTAAATATA 5033
XX

QY 1317 CAATAAAGAGAAAAACATTTTCTGAAAGAGAGAAATAAAGTTTA 1358
DB 5032 TCAAAATATAATATTATTAAATAAATAAATAAATAAATAA 4991

RESULT 14
ABQ66998/c
ID ABQ66998 standard; DNA; 37515 BP.

AC ABQ66998;
XX
XX
DT 28-AUG-2002 (first entry)
XX

DE Human angiogenesis associated polynucleotide SEQ ID NO 28.

XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antitubercids;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW anti-rheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.

XX Homo sapiens.
OS
XX WO200246454-A2.

XX 13-JUN-2002.

XX 06-DEC-2001; 2001WO-EP14320.

XX 06-DEC-2000; 2000DE-1061338.

XX (EPIG-) EPIGENOMICS AG.

XX Schacht O;

XX WPI; 2002-500450/53.

XX New nucleic acid fragments from chemically treated
PT angiogenesis-associated genes, useful for determining methylation
PT status, e.g. in diagnosis or treatment of cancer -

XX Claim 1; SEQ ID NO 28; 41pp + Sequence Listing; German.

XX The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (AB666971-AB667178) or their complements (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 37515 BP; 11428 A; 425 C; 7165 G; 18497 T; 0 other;

Query Match 2.7%; Score 45.6; DB 24; Length 37515;
Best Local Similarity 48.8%; Pred. No. 1.8; Indels 0; Gaps 0;
Matches 123; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 1124 CCTAACCAATAAAAAATTTTGAAGATCTTCCACACATAGAAAAGTTAAATTTGATCAGC 1183
DB 15772 CATACAAAAAAATTTACTTAAATAAATAAATAAATAAATAAATCCCACTCAATATTA 15713

QY 1184 GATGGAATTTTGTACAAAGCTAGTATTTTCATTTGGAGTGCTACTAGTACTAGTAG 1243

DB 15712 AAACAAACATTTAAACATTTACTTAATCTAATTTACTAATAATATAAATAAATATTATT 15653

QY 1244 TACTACCAAGATGAGTTTCTGATTTGGATTTTGAAGCTTTTCTTAGGTTTAAAAACAA 1303

DB 15652 TACCTTTTAAATAACAATTAATACTAAAAAACTAAAAAAATTTATCTTAATATAAAAAA 15593
QY 1304 GTATATTACTAAACAATTAATAAAGAAAAACATTTTGTGAAAAGAGAAATAAAGTTTACTGGA 1363
DB 15592 ACATTTTCTTAATATAAATAACAATAACAATTTTAACTTAAAAAAATAATATTATTAA 15533
QY 1364 CCCATTGTACA 1375
DB 15532 ATCTTCTCTATA 15521

RESULT 15
ABL70563/c
ID ABL70563 standard; DNA; 6352 BP.

XX ABL70563;

XX AC
XX 01-JUL-2002 (first entry)
XX

XX Chemically treated cell signalling DNA sequence#227.

XX Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.

XX Unidentified.

XX WO200202807-A2.

XX 10-JAN-2002.

XX 29-JUN-2001; 2001WO-EP07471.

XX 30-JUN-2000; 2000DE-1032529

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A. Piepenbrock C. Berlin K.

XX WPI; 2002-154758/20.

XX Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling -

XX Claim 1; SEQ ID NO 453; 24pp+sequence listing; English.

XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.

CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.

XX Sequence 6352 BP; 1743 A; 109 C; 1264 G; 3236 T; 0 other;

Query Match 2.7%; Score 45.4; DB 24; Length 6352;

Best Local Similarity 51.4%; Pred. No. 1.3;

Matches 131; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

QY 1104 TTAAGTAAATCTTTTTCGCTAACCAATAAATAATTTTGAAA---ATCTTTCACCA 1160

DB 294 TTAATAAATAATCATATCATCTACTTATATAAATAAATAAATAAATTCATAAACAACA 235

```

1161 TAGAAAGTTAAATTTTCATCAGCGATGGAATTTTGTACAAAGCTAGGTATTTTCATTG 1220
234 AAAAAAAAAACATTAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 175
1221 GGAGTGTACTAGTAAGTACTAACCAGATGAGTTTCTGATTTTGGATTTTGAA 1280
174 TCTCTAACTTTATCAAACTCTATTAATAAATAATTTTAAATTTCTCTAATATTTCTAAA 115
1281 GCCTTTCTTAGGTTAAAAACAAGTATATTACTAAACAATAAAGAAAAACATTTTGTGA 1340
114 CCTATCTACTAAAAAATAATTTTATTAATAAACAATAAATAAATAAATAAATAAATA 55
1341 AAAGAGAAATAAGT 1355
54 AAAAAAAAAATAATT 40

```

Search completed: February 1, 2004, 21:58:56
Job time : 496.261 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:43:46 ; Search time 3880.05 Seconds
(without alignments)
10504.671 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 ggtaagcgttttactatg.....ttctctcagctatatttta 1677

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

- 1: em_estba:*
- 2: em_esthm:*
- 3: em_esthm:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_man:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	295.4	17.6	317	29	BZ662927 SALK_0264
2	246.8	14.7	267	29	AL761108 Arabidops
3	233.2	13.9	278	29	BZ377781 SALK_1061
4	233	13.9	347	9	AV521636 AV521636

C	5	164.6	9.8	232	28	BH612074
C	6	164.6	9.8	233	28	BH612075
	7	101	6.0	486	9	AV825375
	8	86.8	5.2	378	10	BZ523004
	9	82	4.9	146	28	BH617024
C	10	74.2	4.4	825	29	BH493162
C	11	64.6	3.9	735	29	BZ430588
C	12	63.8	3.8	1201	13	BX364179
C	13	63.4	3.8	600	29	CC459772
C	14	62.6	3.7	1201	13	BX335665
C	15	59.4	3.5	936	13	BX404443
C	16	59	3.5	928	29	CNS00DKY
C	17	59	3.5	994	13	BX380865
C	18	58.8	3.5	994	13	BX414650
C	19	57.4	3.4	1101	29	CNS0039G
C	20	57.4	3.4	1101	29	CNS0039Q
C	21	57	3.4	839	28	AQ992711
C	22	57	3.4	1183	13	BX417656
C	23	56.2	3.4	932	9	AL514901
C	24	56	3.3	771	13	BX374423
C	25	56	3.3	1201	13	BX356851
C	26	55.4	3.3	1200	13	BX437758
C	27	55	3.3	1056	13	BX415058
C	28	55	3.3	1167	29	CNS07360
C	29	55	3.3	1200	13	BX437758
C	30	54.8	3.3	1101	29	CNS00EPO
C	31	54.6	3.3	1064	13	BX361825
C	32	54.6	3.3	1201	13	BX399692
C	33	54.2	3.2	1101	29	CNS0022U
C	34	54.2	3.2	1200	29	CNS016CO
C	35	54.2	3.2	1225	29	CNS0161D
C	36	53.6	3.2	1133	13	BX444099
C	37	53.4	3.2	1200	13	BX447469
C	38	53.2	3.2	1002	29	CNS0163Y
C	39	53.2	3.2	1056	13	BX415058
C	40	53.2	3.2	1201	13	BX420717
C	41	53	3.2	1101	29	CNS000B8
C	42	53	3.2	1200	13	BX437750
C	43	53	3.2	1201	9	AL532464
C	44	53	3.2	1201	13	BX426629
C	45	52.6	3.1	1201	9	AL536104

ALIGNMENTS

RESULT 1
BZ662927
LOCUS BZ662927 317 bp DNA linear GSS 31-JAN-2003
DEFINITION SALK_026421.45.45.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_026421.45.45.x, genomic
survey sequence.

ACCESSION BZ662927
VERSION BZ662927.1 GI:28177691
KEYWORDS GSS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 317)
AUTHORS Alonso,J.M., Laissé,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
, C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
, Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL Unpublished
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

Email: eckers@sal.k.edu

This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of AT2g45680.

Class: TDNA tagged.

Location/Qualifiers

```
1..317
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_026421.45.45.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.sal.k.edu/tdna_protocols.html"
BASE COUNT      86 a  75 c  51 g  105 t
ORIGIN
```

```
Query Match      17.6%; Score 295.4; DB 29; Length 317;
Best Local Similarity 99.7%; Pred. No. 1.5e-38;
Matches 296; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1381 TCCGATATATACATAGAGATAGAGCATGAGAGTCAATTTGTTACGTTGATCAA 1440
DB 1 TCCGATATATACATAGAGATAGAGCATGAGAGTCAATTTGTTACGTTGATCAA 60

QY 1441 TCGGAATGTTCTTTAAAGCTCATCGACACATCAGGACCGTTGATTTTCCCGATCAA 1500
DB 61 TCGGAATGTTCTTTAAAGCTCATCGACACATCAGGACCGTTGATTTTCCCGATCAA 120

QY 1501 AAAGCGTTGAATATATCTCAGTTTCTTCTGCTCTATATATATATCTGACGAGTC 1560
DB 121 AAAGCGTTGAATATATCTCAGTTTCTTCTGCTCTATATATATATCTGACGAGTC 180

QY 1561 ACATTTAGTAATCTCTTGGAGTGTACGCGGTAAACGATTTCTTCCATTTGATCC 1620
DB 181 ACATTTAGTAATCTCTTGGAGTGTACGCGGTAAACGATTTCTTCCATTTGATCC 240

QY 1621 GCTTTTAAACACTCTCGTCGTCATCCACCGTCGGTTTCTCTCAGCATATATTTTA 1677
DB 241 GCTTTTAAACACTCTCGTCGTCATCCACCGTCGGTTTCTCTCAGCATATATTTTA 297
```

```
RESULT 2
AL761108
LOCUS
DEFINITION
Arabidopsis thaliana T-DNA flanking sequence GK-205F06-014511,
genomic survey sequence.
ACCESSION
AL761108
VERSION
AL761108.1 GI:21502149
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
```

```
REFERENCE
1 Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saedler,H.
and Weisshaar,B.
A pipeline for automated high-throughput generation of FSTs
(flaning sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
Unpublished
JOURNAL
REFERENCE
Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
for flanking sequence tag based reverse genetics
Unpublished
JOURNAL
REFERENCE
3 (bases 1 to 267)
```

AUTHORS

Direct Submission

JOURNAL

COMMENT

FEATURES

Source

```
1..267
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-205F06-014511"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"
BASE COUNT      64 a  48 c  68 g  87 t
ORIGIN
```

```
Query Match      14.7%; Score 246.8; DB 29; Length 267;
Best Local Similarity 99.2%; Pred. No. 1.4e-30;
Matches 248; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 385 GCAGAAAGTTGGTGGCGTGAAGAAATTAACAAATTAATCTCTGCTGACCGCAATCTGT 444
DB 18 GCGAGAGTTGGTGGCGTGAAGAAATTAACAAATTAATCTCTGCTGACCGCAATCTGT 77

QY 445 AATCGGAAAGTGGAAACCCACTTGGTTAACTTTAAAGCCCTAAAGGTTACTACCGGTT 504
DB 78 AATCGGAAAGTGGAAACCCACTTGGTTAACTTTAAAGCCCTAAAGGTTACTACCGGTT 137

QY 505 TCACCGGTTTATAATTTGGTGTAAATTTCTAATCCCGATCCGTTTGTGTTTAACTCTCA 564
DB 138 TCACCGGTTTATAATTTGGTGTAAATTTCTAATCCCGATCCGTTTGTGTTTAACTCTCA 197

QY 565 AGGCCACGTTATCGCCCAATATTTGATTTTGGTGGGTAGGAATGTTGGGTCGAATA 624
DB 198 AGGCCACGTTATCGCCCAATATTTGATTTTGGTGGGTAGGAATGTTGGGTCGAATA 257

QY 625 GTTGGGCGCTA 634
DB 258 GTTGGGCGCTA 267
```

RESULT 3

BZ377781

LOCUS

DEFINITION

SALK_106185.39.30.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_106185.39.30.x, genomic
survey sequence.

ACCESSION

BZ377781

VERSION

BZ377781.1 GI:25467878

KEYWORDS

GSS.

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (bases 1 to 278)

AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab

,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.

TITLE Zimmerman, J. and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Arabidopsis Genome
COMMENT Unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At2g45680.

FEATURES
source

Class: TDNA tagged.
Location/Qualifiers
1. .278

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_106185.39.30.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html."

BASE COUNT 72 a 69 c 44 g 93 t

ORIGIN

Query Match 13.9%; Score 233.2; DB 29; Length 278;
Best Local Similarity 96.5%; Pred.No. 2.3e-28;
Matches 249; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

1421 GATTGTTTCACGGTGTACATCGGA-ATGGTTCCTTTTAAAGCTCATCGACACATCAGGAC 1479

1. GATTTCATCAGCTGTGTACATCGGATCGATGGATCTTTATAGCTCATCCACACATCAAGAG 60

1480 CGTTGATTTTCCCGCATCAAAAAGGTTGATATCTCTCACTGTTTCTCTCTCT 1539

61 CGTTGATTTTCCCGCATCAAAAAGGTTGATATCTCTCACTGTTTCTCTCTCT 120

1540 ATATATATATCTCTGACGAGTCACATTTAGTATCTCTTGGACGTGTAAACGCGTTAAAA 1599

121 ATATATATATCTCTGACGAGTCACATTTAGTATCTCTTGGACGTGTAAACGCGTTAAAA 180

1600 CGATTCTTCCCATGATCCGGTTTAAACACTCTCTGTCATCTCCACCGTCCGTTT 1659

181 CGATTCTTCCCATGATCCGGTTTAAACACTCTCTGTCATCTCCACCGTCCGTTT 240

1660 TCTCTCAGCTATATTTTA 1677

241 TCTCTCAGCTATATTTTA 258

RESULT 4

AV521636/c

LOCUS 347 bp mRNA linear EST 07-SEP-2000

DEFINITION AV521636 Arabidopsis thaliana aboveground organs two to six-week

AV521636 Old Arabidopsis thaliana cDNA clone AP263b12f 3', mRNA sequence.

AV521636 AV521636.1 GI:86811163

EST.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 347)

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation

of 12,028 non-redundant expressed sequence tags from normalized and

size-selected cDNA libraries

DNA Res. 7, 175-180 (2000)

20363093

10907847

COMMENT Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 252-0812, Japan

Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

Location/Qualifiers

1. .347

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="AP263b12f"

/tissue_type="aboveground organs"

/dev_stage="two to six-week old"

/clone_lib="Arabidopsis thaliana aboveground organs two to

six-week old"

/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:

XhoI"

BASE COUNT 102 a 77 c 54 g 114 t

ORIGIN

Query Match 13.9%; Score 233; DB 9; Length 347;

Best Local Similarity 97.9%; Pred.No. 2.4e-28;

Matches 236; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATTCATTTGTTGAATGC 60

241 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATTCATTTGTTGAATGC 182

QY 61 TTTTTCAGATCATCAAGGCTCTCAGATTTCTTAGGGAATGTTTCAGGCTTTGTTA 120

181 TTTTTCAGATCATCAAGGCTCTCAGATTTCTTAGGGAATGTTTCAGGCTTTGTTA 122

QY 121 GAAATGTTGTTTATGCAACAGGTAGAGCAATACCATAGACAGATGTATCTGAAGAGA 180

121 GAAATGTTGTTTATGCAACAGGTAGAGCAATACCATAGACAGATGTATCTGAAGAGA 62

QY 181 TAGCTTCTCTATGTCTAAAGAAATGGACCGATACGAATAAACAAGCATCATAAAGAT 240

61 TAGCTTCTCTATGTCTAAAGAAATGGACCGATACGAATAAACAAGCATCATAAAGAT 2

QY 241 T 241

1 T 1

RESULT 5

BH612074/c

LOCUS 232 bp DNA linear GSS 04-JAN-2002

DEFINITION SALK_032102 Arabidopsis thaliana TDNA insertion lines Arabidopsis

thaliana genomic clone SALK_032102, genomic survey sequence.

ACCESSION BH612074

VERSION BH612074.1

GI:18059525

GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 232)

Alonso, J.M., Leisner, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab

, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.

, Zimmerman, J., and Ecker, J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.

Location/Qualifiers

FEATURES

source

1. .232
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_032102"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"
59 a 45 c 46 g 82 t

BASE COUNT
ORIGIN

Query Match 9.8%; Score 164.6; DB 28; Length 232;
Best Local Similarity 97.7%; Pred. No. 4e-17; Indels 0; Gaps 0;
Matches 167; Conservative 0; Mismatches 4;
QY 183 AGCTTCTCTATGCTTAAGAAATGGACCGATAGCAATATAAACAAGCATCATTAAAGATTAA 242
Db 188 AGCTTCTCTATGCTTAAGAAATGGACCGATAGCAATATAAACAAGCATCATTAAAGATTAA 129
QY 243 AATGGTTTGTAAAGAAATACACCTTATTATGTGAATTTGTGGTTAGTGAAGATTAA 302
Db 128 AATGGTTTGTAAAGAAATACACCTTATTATGTGAATTTGTGGTTAGTGAAGATTAA 69
QY 303 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTATGAT 353
Db 68 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTCGTAA 18

RESULT 6

BH612075/c

LOCUS

DEFINITION BH612075 233 bp DNA linear GSS 04-JAN-2002
SALK_032104 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_032104, genomic survey sequence.

ACCESSION BH612075

VERSION BH612075.1

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 233)

AUTHORS Alonso, J.M., Leisse, T.J., Baxajae, P., Chen, H., Cheuk, R., Gadkinab

, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,

Zimmerman, J., and Ecker, J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished

CONTACT: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA.

Class: TDNA tagged.

Location/Qualifiers

FEATURES

source

1. .232
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_032104"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"
58 a 45 c 47 g 83 t

source

1. .233
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_032104"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"
58 a 45 c 47 g 83 t

BASE COUNT
ORIGIN

Query Match 9.8%; Score 164.6; DB 28; Length 233;
Best Local Similarity 97.7%; Pred. No. 4e-17; Indels 0; Gaps 0;
Matches 167; Conservative 0; Mismatches 4;
QY 183 AGCTTCTCTATGCTTAAGAAATGGACCGATAGCAATATAAACAAGCATCATTAAAGATTAA 242
Db 188 AGCTTCTCTATGCTTAAGAAATGGACCGATAGCAATATAAACAAGCATCATTAAAGATTAA 129
QY 243 AATGGTTTGTAAAGAAATACACCTTATTATGTGAATTTGTGGTTAGTGAAGATTAA 302
Db 128 AATGGTTTGTAAAGAAATACACCTTATTATGTGAATTTGTGGTTAGTGAAGATTAA 69
QY 303 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTATGAT 353
Db 68 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTCGTAA 18

RESULT 7

AV825375

LOCUS

DEFINITION AV825375 RAF17 Arabidopsis thaliana cDNA clone RAFL07-08-P04 5',
mRNA sequence.

ACCESSION AV825375

VERSION AV825375.1

KEYWORDS EST.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 486)

AUTHORS Seki, M., Narusaka, M., Ishida, J., Kaniya, A., Satou, M., Nakajima, M.,

Cono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,

Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,

and Shinozaki, K.

Large scale analysis of Arabidopsis full-length cDNA (2002b)

Unpublished

CONTACT: Motoaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: mseki@ctc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially

as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

and XhoI was ligated to modified Lambda FIC-1 vector (Garnick et

al., submitted for publication) digested with BamHI and SalI. This

clone is in a modified pBluescript vector. Please visit our web

site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further

details.

Location/Qualifiers

1. .486

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/clone="RAFL07-08-P04"

```

/dev stage="rosette plants"
/lab_host="Dhios"
/clone_lib="RAFL7"
/note="Site 1: BamHI; Site 2: SalI; subjected to
cold-treated (1, 2, 5, 10, 24 hr)"
BASE COUNT      131 a      117 g      117 t      3 others
ORIGIN
Query Match      6.0%; Score 101; DB 9; Length 486;
Best Local Similarity 100.0%; Pred. No. 8.2e-07;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1577 TTGACGTGAACGCGTAAACAGATTCTTCCCATTTGATCCGCTTTTAAACAATCTCTGTCATCTC 1636
Db 1 TTGACGTGAACGCGTAAACAGATTCTTCCCATTTGATCCGCTTTTAAACAATCTCTGTCATCTC 60

2Y 1637 GTGCTATCTCCACCGTCGCTTTCTCTCAGCTATATTTTA 1677
Db 61 GTGCTATCTCCACCGTCGCTTTCTCTCAGCTATATTTTA 101

RESULT 8
3E23004
LOCUS
DEFINITION M31B6STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION BE523004
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE
AUTHORS White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilardeya, C., Jaworski, J.G., Ohlrogge, J. and Benning, C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
PUBMED 11115876
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371

FEATURES
source
1..378
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M31B6"
/tissue_type="seed"
/dev stage="5-13 days after flowering"
/lab host="E.coli"
/clone_lib="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site 1: EcoRI; Site 2: XhoI"
BASE COUNT      108 a      97 c      81 g      92 t
ORIGIN
Query Match      5.2%; Score 86.8; DB 10; Length 378;
Best Local Similarity 97.8%; Pred. No. 0.00018;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1588 ACCCGTTAAACAGATTCTTCCCATTTGATCCGCTTTTAAACAATCTCTGTCATCTC 1647
Db 13 ACAGAGTTAAACAGATTCTTCCCATTTGATCCGCTTTTAAACAATCTCTGTCATCTC 72

QY 1648 CACGTCGCTTTCTCTCAGCTATATTTTA 1677
Db 73 CACGTCGCTTTCTCTCAGCTATATTTTA 102

RESULT 9
BH617024
LOCUS
DEFINITION SALK_035853 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_035853, genomic survey sequence.
ACCESSION BH617024
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE
AUTHORS Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadzinab
, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,
Zimmerman, J. and Ecker, J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL Unpublished
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within 300 bases of the 5' end of
At2g45680.
Class: TDNA tagged.
Location/Qualifiers
1..146
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_035853"
/note="lib="Arabidopsis thaliana TDNA insertion lines"
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at: http://signal.salk.edu/tdna_protocols.html"
BASE COUNT      33 a      44 c      27 g      42 t
ORIGIN
Query Match      4.9%; Score 82; DB 28; Length 146;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1596 AAAACGATTCTTCCCATTTGATCCGCTTTTAAACAATCTCTGTCATCTC 1655
Db 1 AAAACGATTCTTCCCATTTGATCCGCTTTTAAACAATCTCTGTCATCTC 60

QY 1656 GTTTCTCTCAGCTATATTTTA 1677
Db 61 GTTTCTCTCAGCTATATTTTA 82

RESULT 10
BH493162/c

```


This is single pass sequence recovered from the left border of TDNA.

Location/Qualifiers
1. .600

```

/organism="Arabidopsis thaliana"
/mol type="genomic DNA"

```

`</clone>SALK_13376.34.05.x"`
`</clone_lib>Arabidopsis thaliana TMDA insertion lines"`
`/notes>`PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TMDA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used
be found at http://signal.salk.edu/tmda_protocols.htm
130 c 102 g 166 t 42 others
</text></?>

each of which contains one or more TDNA insertion elements. The resultant fragment for each library was performed on a 1% agarose gel.

directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/tdna_protocols.html

Category	Score	DB	Length
3.8%	63.4	29	600
81.3%	1.1		

[illegible]

416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927

TTTGTGCCTAACCAATAAATAATTATGAAAATCTTTCCAACC 1159
GGGCTTCAACCCCAAGATATGGCTTACAGCTTTCTCATATTAATAGTATGAA 1160

TTTAGCGTATCCAAATAAAAATTATGGAAAAATCTTTACCAAC 369

100

Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
1201 bp mRNA linear EST 01-MAY-2003

ODI017YJ04 5-PRIME, mRNA sequence.

1 GI:30308396

tens (human)
tens
Chordata: Craniata: Vertebrata: Euteleostomi:
Metazoa:

1; Metazoa; Chordata; Craniata; Vertebrata; Mollusca; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 to 1201)

Gruber, C., Jessee, J. and Polayes, D.
with cDNA libraries and normalization

Genoscope

e - Centre National de Sequencage
1006 EVRY cedex - France

eqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
The database was constructed by Life Technologies, a division of

an. This sequence belongs to sequence cluster 7077.f
: Feng Liang Email: fliang@lifetech.com URL :
: www.lifetech.com Generation: 1600

length.invitrogen.com/ Invitrogen Corporation 1600
avenue Genoscope sequence ID : CS0DI017DE02QP1.
Location/Qualifiers

Location/Qualifiers
1. .1201
/organism "Uomo sapiens"

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:45:16 ; Search time 110.457 Seconds
(without alignments)
6701.220 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 99ttagcggtttactatg.....tttctctcagatatatttta 1677

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	46.4	2.8	7218	1	US-08-332-463-14
C 2	42.2	2.5	1666	1	US-08-076-090-1
C 3	42.2	2.5	1666	5	PCT-US94-06661-1
C 4	42.2	2.5	2500	1	US-08-550-715-10
C 5	42.2	2.5	3385	4	US-09-411-449-3
C 6	42.2	2.5	3496	4	US-09-411-449-1
C 7	42.2	2.5	3805	4	US-08-411-449-4
C 8	42.2	2.5	3916	4	US-09-411-449-2
C 9	42.2	2.5	6060	5	PCT-US96-09430-7
C 10	42	2.5	1440	4	US-09-107-532A-2589
C 11	39.6	2.4	837	3	US-08-998-416-288
C 12	39.6	2.4	7218	1	US-08-232-463-14
C 13	39	2.3	721	1	US-08-451-405A-2
C 14	39	2.3	2251	3	US-08-991-677-11
C 15	39	2.3	1664976	4	US-08-916-421B-1
C 16	39	2.3	1664976	4	US-08-916-421B-1
C 17	38.8	2.3	2169	4	US-09-434-408-3
C 18	38.6	2.3	606	4	US-09-601-198-162
C 19	38.6	2.3	1990	4	US-08-961-527-232
C 20	38.6	2.3	19124	2	US-08-487-826B-13
C 21	38.2	2.3	731	1	US-08-451-405A-2
C 22	38.2	2.3	19250	4	US-08-961-527-35
C 23	38	2.3	2341	3	US-09-187-049-11
C 24	37.8	2.3	1189	1	US-08-307-591-2
C 25	37.6	2.2	2394	3	US-09-414-010-3
C 26	37.6	2.2	8133	1	US-08-480-604A-5
C 27	37.6	2.2	8133	2	US-08-405-496A-5

Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 1, Appli
Sequence 25, Appli
Sequence 25, Appli
Sequence 25, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 56, Appli
Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHREIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Poley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION: 435
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s

US-08-232-463-14 Query Match 2.8%; Score 46.4; DB 1; Length 7218;


```
Best Local Similarity 12.9%; Pred. No. 0.019;
Matches 59; Conservative 189; Mismatches 210; Indels 0; Gaps 0;
QY 138 AACAGGTAGAGACATACCATAGACATATCTGAGAGATAAGCTTCTCTATGCT 197
DB 1398 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1329
QY 198 AAAGAAATGGACGATACGAAATAAAACAAAGCATCAATTAAGATTAAATGGTTGTAGAA 257
DB 1328 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1269
QY 258 ATACTACACTTATTTATGTGAAATGTGTGTAGTGAAGTAAAGTAAACATCGGAATCCA 317
DB 1268 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1209
QY 318 AAACCTCAATTTACCAATCAGCCCAATTAATGATGCTGCGTAATGAATGGTATGCTG 377
DB 1208 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1149
QY 378 ATGTAGGCAAAAGTTGTGCTGCGAAATAACAATATATCCCTCTGTGGTGACCG 437
DB 1148 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1089
QY 438 AATCTGTAATCGGAAAGCTGGAACCACTGGTTAACTTTTAAGCTTAAAGTTACT 497
DB 1088 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1029
QY 498 ACCGGTTGACCGTTTATAATTTGGTTTAAATCTTAATCCCGATCCGTTTGTGTT 557
DB 1028 AATCTGTGAGCGTATGCAACGAAGGAAATAGTTATAGTAGCGCACTCGATGGGA 969
QY 558 AATCTCAGGCGCAGCTTATCGCAATATTTGATTTT 595
DB 968 CATTTCAACGTAACCGTTTAAATAATTTGATCTT 931

RESULT 2
US-08-076-090-1
; Sequence 1, Application US/08076090
; Patent No. 5631162
; APPLICANT: LeBouch, Philippe
; APPLICANT: London, Irving M.
; APPLICANT: Tuan, Dorothy
; TITLE OF INVENTION: Retroviral Vectors for Transducing
; TITLE OF INVENTION: Beta-Globulin Gene and Beta-Locus Control Region
; TITLE OF INVENTION: Derivatives
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/076.090
; FILING DATE: 19930611
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MIT 6128
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 815-6508
; TELEFAX: (404) 815-6555
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: Beta-globin gene
; FEATURE:
; NAME/KEY: misc signal
; LOCATION: 37..298
; OTHER INFORMATION: /note= "Exon III"
; FEATURE:
; NAME/KEY: misc signal
; LOCATION: 299..1148
; OTHER INFORMATION: /note= "Intron 2"
; FEATURE:
; NAME/KEY: misc signal
; LOCATION: 1149..1370
; OTHER INFORMATION: /note= "Exon II"
; FEATURE:
; NAME/KEY: misc signal
; LOCATION: 1371..1501
; OTHER INFORMATION: /note= "Intron 1"
; FEATURE:
; NAME/KEY: misc signal
; LOCATION: 1502..1643
; OTHER INFORMATION: /note= "Exon I"
; US-08-076-090-1
Query Match 2.5%; Score 42.2; DB 1; Length 1666;
Best Local Similarity 47.5%; Pred. No. 0.14;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 1091 TTTTCCATAAAATTAAGTAAATCTTTTTCCTAACCAATAAAATTAATTCGAAATC 1150
DB 684 TGTGTACACATATTAACATACATTTAACCCATTAATATGTAATGTTATGTAATC 743
QY 1151 TTTCACACATAGAAAGTTAAATTTGATCAGCATGGAATTTTGTACAAAGCTAGGT 1210
DB 744 AATTCGAAATTAAGAAATTAAGTAGGAGATTATGAATATGCAATTAAGCACATAT 803
QY 1211 ATTTTCATTTGGGAGTGTACTAGTACTAGTAACTAACCCAGATGAGTTTCTGATTTT 1270
DB 804 ATTCGAAATAGTAATGTAAGTACTAGGAGACTGTGTAAAGTTTCTTAAAGTTTACTTAATGT 863
QY 1271 GGATTTTGAAGCTTTTCTTAGGTTTAAACAAAGTATATTTACTAAACAAATAAAAGAAAA 1330
DB 864 ATCTCAGAGATATTTCTTTTGTATACAAATGTTAAGGCATTAAAGTATAATAGTAAAA 923
QY 1331 CATTTTGTGAAAGAGAAATAAA 1353
DB 924 ATTCGGAGAGAAAGAAAAAGA 946

RESULT 3
PCT-US94-06661-1
; Sequence 1, Application PC/TUS9406661
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Retroviral Vectors for Transducing
; TITLE OF INVENTION: Beta-Globulin Gene and Beta-Locus Control Region
; TITLE OF INVENTION: Derivatives
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06661
```

FILING DATE: 10-JUN-1994
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1666 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
CELL TYPE: Beta-globin gene
FEATURE:
NAME/KEY: misc signal
LOCATION: 37..298
OTHER INFORMATION: /note= "Exon III"
FEATURE:
NAME/KEY: misc signal
LOCATION: 299..1148
OTHER INFORMATION: /note= "Intron 2"
FEATURE:
NAME/KEY: misc signal
LOCATION: 1149..1370
OTHER INFORMATION: /note= "Exon II"
FEATURE:
NAME/KEY: misc signal
LOCATION: 1371..1501
OTHER INFORMATION: /note= "Intron 1"
FEATURE:
NAME/KEY: misc signal
LOCATION: 1502..1643
OTHER INFORMATION: /note= "Exon I"

CT-US94-06661-1

Query Match 2.5%; Score 42.2; DB 5; Length 1666;
Best Local Similarity 47.5%; Pred. No. 0.14;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
Y 1091 TTTTCCATAAATAAAGTAATCTTTTTCCTAACCAATAAAATTTATGAAATC 1150
b 684 TGTGTACACATATTAACCAATTTACCTTTAACCCATTAATATGTAATGATTATGATC 743
Y 1151 TTTCCAAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
b 744 ATTCGAAATAAAGAAATAAAGTAGGGAGATTAATGATGCAATTAACACACATAT 803
Y 1211 ATTCATTGGGAGTGCTACTAGTAAGTAACTAACCCAGATGAGTTTCTGATTTT 1270
b 804 ATTCGAAATAGTAATGCTACTAGGCAGACTGTGTAAGTTTATTAAGTTACTTAATGT 863
Y 1271 GGATTTTGAAGCTTTTCTTAGGTTAAACCAAGATATTAACCAATTAAGAGAAA 1330
b 864 ATTCAGAGATATTTCCCTTTTGTATACACATGTTAAGGCAATTAATTAATAGTAAA 923
Y 1331 CATTTTGTGAAAGAGAAATAAA 1353
b 924 ATTGGGAGAGAGAAAAAAGAGA 946

RESULT 4
US-08-550-715-10/c
Sequence 10, Application US/08550715
Patent No. 5750345
GENERAL INFORMATION:
APPLICANT: Bowie, Lemuel J.
TITLE OF INVENTION: Human -Thalassemia Mutations as a Predictor of
TITLE OF INVENTION: Blood-Related Disorders
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSES: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/550,715
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 28493/32834
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: join(687..778, 909..1131, 1982..2107)
US-08-550-715-10

Query Match 2.5%; Score 42.2; DB 1; Length 2500;
Best Local Similarity 47.5%; Pred. No. 0.16;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 1091 TTTTCCATAAATAAAGTAATCTTTTTCCTAACCAATAAAATTTATGAAATC 1150
Db 1596 TGTGTACACATATTAACCAATTTACCTTTAACCCATTAATGTAATGATTATGATC 1537
QY 1151 TTTCCAAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
Db 1536 AATTAAAAATAAAGAGAAATAAAGTAGGAGATTTATGAATATGCAATAAAGCACACATAT 1477
QY 1211 ATTCATTGGGAGTGCTACTAGTAAGTAACTAACCCAGATGAGTTTCTGATTTT 1270
Db 1476 ATTCGAAATAGTAATGCTACTAGGCAGACTGTGTAAGTTTATTAAGTTACTTAATGT 1417
QY 1271 GGATTTTGAAGCTTTTCTTAGGTTAAACCAAGATATTAACCAATTAAGAGAAA 1330
Db 1416 ATCTCAGAGATATTTCCCTTTTGTATACACATGTTAAGGCAATTAATTAATAGTAAA 1357
QY 1331 CATTTTGTGAAAGAGAAATAAA 1353
Db 1356 ATTGGGAGAGAGAAAAAAGAGA 1334

RESULT 5
US-09-411-449-3/c
Sequence 3, Application US/09411449
Patent No. 6524851
GENERAL INFORMATION:
APPLICANT: James Ellis
TITLE OF INVENTION: HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
TITLE OF INVENTION: Beta-GLOBIN REGULATORY ELEMENTS
FILE REFERENCE: 17860017
CURRENT APPLICATION NUMBER: US/09/411,449
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: 2,246,005
PRIOR FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent In Ver. 2.1

```
; SEQ ID NO 3
; LENGTH: 3385
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-449-3

Query Match      2.5%; Score 42.2; DB 4; Length 3385;
Best Local Similarity 47.5%; Pred. No. 0.18;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1091 TTTTCCATAAAATTAAGTAAATCTTTTGGCTTAACCAATAAAATATTGAAAATC 1150
DB 2128 TGTGTACACATATTAAACATTAACCTTTAACCCATAAATATGATATGATATGATC 2069
QY 1151 TTTCACCAATAGAAAAGTTAAATTTGATCGCGATGGAATTTTGTACAAAGCTAGGT 1210
DB 2068 AATTAAAAATAAAGAAAATAAAGTAGGAGATTATGAATGCAAAATAAGCACACATAT 2009
QY 1211 ATTTCATTGGGAGTGTTACTAGTAAGTAACTAAGTAACTAAGTAACTAAGTAACT 1270
DB 2008 ATTCCAAATAGTAATGTACTAGGCGAGCTGTGTAAAGTTTCTTTTAAAGTTACTTA 1949
QY 1271 GGATTTTGAAGCTTTCTTAGTGTAAATAAACAAGTATATTACTAAACAATAAAGAAA 1330
DB 1948 ATCTCAGAGATATTTCTTTTGTGTATACACATGTTAAGGCAATTAAGTATAATAGTAAA 1889
QY 1331 CATTTTGTGAAAAGAGAAATAAA 1353
DB 1888 ATTGCGGAGAGAAAAGAAAAGAA 1866

RESULT 6
US-09-411-449-1/c
; Sequence 1, Application US/09411449
; Patent No. 6524851
; GENERAL INFORMATION:
; APPLICANT: James Ellis
; TITLE OF INVENTION: HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
; FILE OF INVENTION: beta-GLOBIN REGULATORY ELEMENTS
; CURRENT APPLICATION NUMBER: US/09/411,449
; PRIOR FILING DATE: 1999-10-01
; PRIOR FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-449-1

Query Match      2.5%; Score 42.2; DB 4; Length 3496;
Best Local Similarity 47.5%; Pred. No. 0.18;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1091 TTTTCCATAAAATTAAGTAAATCTTTTGGCTTAACCAATAAAATATTGAAAATC 1150
DB 2239 TGTGTACACATATTAAACATTAACCTTTAACCCATAAATATGATATGATATGATC 2180
QY 1151 TTTCACCAATAGAAAAGTTAAATTTGATCGCGATGGAATTTTGTACAAAGCTAGGT 1210
DB 2179 AATTAAAAATAAAGAAAATAAAGTAGGAGATTATGAATGCAAAATAAGCACACATAT 2120
QY 1211 ATTTCATTGGGAGTGTTACTAGTAAGTAACTAAGTAACTAAGTAACTAAGTAACT 1270
DB 2119 ATTCCAAATAGTAATGTACTAGGCGAGCTGTGTAAAGTTTCTTTTAAAGTTACTTA 2060
QY 1271 GGATTTTGAAGCTTTCTTAGTGTAAATAAACAAGTATATTACTAAACAATAAAGAAA 1330
DB 2059 ATCTCAGAGATATTTCTTTTGTATACACATGTTAAGGCAATTAAGTATAATAGTAAA 2000
QY 1331 CATTTTGTGAAAAGAGAAATAAA 1353
```

```
DB 1999 ATTGCGGAGAGAAAAGAAAAGAA 1977

RESULT 7
US-09-411-449-4/c
; Sequence 2, Application US/09411449
; Patent No. 6524851
; GENERAL INFORMATION:
; APPLICANT: James Ellis
; TITLE OF INVENTION: HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
; FILE OF INVENTION: beta-GLOBIN REGULATORY ELEMENTS
; CURRENT APPLICATION NUMBER: US/09/411,449
; PRIOR FILING DATE: 1999-10-01
; PRIOR FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3805
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-449-4

Query Match      2.5%; Score 42.2; DB 4; Length 3805;
Best Local Similarity 47.5%; Pred. No. 0.19;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1091 TTTTCCATAAAATTAAGTAAATCTTTTGGCTTAACCAATAAAATATTGAAAATC 1150
DB 2548 TGTGTACACATATTAAACATTAACCTTTAACCCATAAATATGATATGATATGATC 2489
QY 1151 TTTCACCAATAGAAAAGTTAAATTTGATCGCGATGGAATTTTGTACAAAGCTAGGT 1210
DB 2488 AATTAAAAATAAAGAAAATAAAGTAGGAGATTATGAATGCAAAATAAGCACACATAT 2429
QY 1211 ATTTCATTGGGAGTGTTACTAGTAAGTAACTAAGTAACTAAGTAACTAAGTAACT 1270
DB 2428 ATTCCAAATAGTAATGTACTAGGCGAGCTGTGTAAAGTTTCTTTTAAAGTTACTTA 2369
QY 1271 GGATTTTGAAGCTTTCTTAGTGTAAATAAACAAGTATATTACTAAACAATAAAGAAA 1330
DB 2368 ATCTCAGAGATATTTCTTTTGTATACACATGTTAAGGCAATTAAGTATAATAGTAAA 2309
QY 1331 CATTTTGTGAAAAGAGAAATAAA 1353
DB 2308 ATTGCGGAGAGAAAAGAAAAGAA 2286

RESULT 8
US-09-411-449-2/c
; Sequence 2, Application US/09411449
; Patent No. 6524851
; GENERAL INFORMATION:
; APPLICANT: James Ellis
; TITLE OF INVENTION: HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
; FILE OF INVENTION: beta-GLOBIN REGULATORY ELEMENTS
; CURRENT APPLICATION NUMBER: US/09/411,449
; PRIOR FILING DATE: 1999-10-01
; PRIOR FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3916
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-449-2

Query Match      2.5%; Score 42.2; DB 4; Length 3916;
Best Local Similarity 47.5%; Pred. No. 0.19;
```

Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

1091 TTTTTCATAAAATTAAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATC 1150
 2659 TGTGTACACATATTAAACATTACACTTTAACCCATAATATGTAATGATTATGTAATC 2600
 1151 TTTCCAAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
 2599 AATTAAATAAAGAAAAATAAGTAGGAGATTATGAATATGCAATAAGCACACATAT 2540
 1211 ATTTTCATTTGGGAGTCTACTAGTAACTAGTAACTAACCAGATGAGTTCTGATTTT 1270
 2539 ATTCCAAATAGTAATGTAAGTACGACACTGTGTAAAGTTTTTTTAAAGTTACTTAATGT 2480
 1271 GGATTTTGAAGCTTTTCTTTAGGTTAAAAACAGTATATTAACCAATAAAAGAAAAA 1330
 2479 ATCTCAGAGATATTTCTTTTGTATACACAATGTTAAGGCATTAAAGTATATAGTAAAA 2420
 1331 CATTTTGTGAAGAGAAATATA 1353
 2419 ATTGCGGAGAGAAAAAAGA 2397

RESULT 9
 CT-US96-09430-7/c
 Sequence 7, Application PC/TUS9609430
 GENERAL INFORMATION:
 APPLICANT: Glazer, Peter M.
 TITLE OF INVENTION: TREATMENT OF HEMOGLOBINOPATHIES
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OncorPharm, Inc.
 STREET: 200 Perry Parkway
 CITY: Gaithersburg
 STATE: Maryland
 COUNTRY: US
 ZIP: 20877
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/09430
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/473,845
 FILING DATE: 07-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Karta, Glenn E.
 REGISTRATION NUMBER: 30,649
 REFERENCE/DOCKET NUMBER: PA-0040
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-527-2058
 TELEFAX: 301-208-6997
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6060 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT: 11
 CT-US96-09430-7

Query Match 2.58; Score 42.2; DB 5; Length 6060;
 Best Local Similarity 47.58; Pred. No. 0.22;

Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

1091 TTTTTCATAAAATTAAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATC 1150
 3096 TGTGTACACATATTAAACATTACACTTTAACCCATAATATGTAATGATTATGTAATC 3037
 1151 TTTCCAAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
 3036 AATTAAATAAAGAAAAATAAGTAGGAGATTATGAATATGCAATAAGCACACATAT 2977
 1211 ATTTTCATTTGGGAGTCTACTAGTAACTAGTAACTAACCAGATGAGTTCTGATTTT 1270
 2976 ATTCCAAATAGTAATGTAAGTACGACACTGTGTAAAGTTTTTTTAAAGTTACTTAATGT 2917
 1271 GGATTTTGAAGCTTTTCTTTAGGTTAAAAACAGTATATTAACCAATAAAAGAAAAA 1330
 2916 ATCTCAGAGATATTTCTTTTGTATACACAATGTTAAGGCATTAAAGTATATAGTAAAA 2857
 1331 CATTTTGTGAAGAGAAATATA 1353
 2856 ATTGCGGAGAGAAAAAAGA 2834

RESULT 10
 US-09-107-532A-2589
 Sequence 2589, Application US/09107532A
 Patent No. 6583275
 GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107,532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/085,598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Arinello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 2589:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1440 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1...1440

SEQUENCE DESCRIPTION: SEQ ID NO: 2589;
US-09-107-532A-2589

Query Match 2.5%; Score 42; DB 4; Length 1440;
Best Local Similarity 54.5%; Pred. No. 0.15;
Matches 84; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 1297 AAACAAGTATATTACTAACAATAAAGAAACATTTTGTGAAAGAGAAATAAGTT 1356
DB 102 AAGAAATCTTAAACCATCAAGCAAAATTCGAACACAAAGGTGTAAAAACAGAAATCAGTT 161
QY 1357 TACTGGACCCCATTTGTACAGATGTCCTCAATAATACTGATAGAGATAGACCAATGGA 1416
DB 162 ATCTGAACAAGAAATCAAGAGGCGAGATGCAATCTTCTTCGAGTAGATAAAGAAATCGA 221
QY 1417 AAGTCATTGTTCAGTGTGTCACAAATCGAATGGT 1450
DB 222 ATTGGACCGATTTCGCGGAAAAAAGTAAGCGT 255

RESULT 11
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgens
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
US-08-998-416-288

Query Match 2.4%; Score 39.6; DB 3; Length 837;

Best Local Similarity 45.7%; Pred. No. 0.53;
Matches 138; Conservative 0; Mismatches 164; Indels 0; Gaps 0;
QY 1091 TTTTTCATATAAATAAAGTAAATCTTTTTTTCCTTAACCAATAAATAAATTTTGAATAATC 1150
DB 601 TTTATTAATTAATTAATAATGATAAATAATTAATTAATAATAATTAATTAATAATAA 542
QY 1151 TTTCCACCATAGAAAAGTTAAATTTTCATCAGCGATCGGAATTTTGTACCAAGCTAGGT 1210
DB 541 TAAATAATAAATAAATGAATATAATTTTAAATAATAATAAATAAATAATAATAAATA 482
QY 1211 ATTTCATTGGGAGTGACTAGTAACCTAGTAAGTACTTAACCAAGATGAGTTTCTGATTTT 1270
DB 481 AAGTTAAATTAATTTTAAATAATAATTTCTTATTAATAAGATTAAATAATAATCAACAT 422
QY 1271 GGATTTTGAACCTTTTCTTAGGTTAAAAAACAAGTATATTTACTTAACAATAAAGAAAA 1330
DB 421 AATATTTTATAAAAAATAGATATATTAATAATAAATAAATAAATAAATAAATAAATA 362
QY 1331 CATTTTGTGAAAAGAGAAATAAAGTTTACTGGCCCATTTGTACAGATGTCCTCCATAATA 1390
DB 361 AATCTTTATAATAATAAATAATAATATTTTAAATAACAATAATAATAATAATAATAAT 302
QY 1391 AT 1392
DB 301 AT 300

RESULT 12
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, P.
; APPLICANT: SCHEIFLINGER, P.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234514)..(234514)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature

LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855399)..(855399)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130981)..(1130981)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DN nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:42:55 ; Search time 630.401 Seconds
(without alignments)
9695.297 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 GGTAAAGCGTTTACTATG.....TTTCTCTAGTATATTTA 1677

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1677	100.0	1677	10	US-09-938-842A-3729
2	1677	100.0	1677	12	Sequence 3729, Ap
3	52.4	3.1	2000	10	US-09-938-842A-3729
4	49.6	3.0	858	15	Sequence 73, Appl
5	49.2	2.9	3673778	13	US-10-198-846-7035
6	49	2.9	3673778	13	US-10-312-841-2
7	47.8	2.9	2958	10	US-09-938-842A-338
8	47.8	2.9	2958	12	Sequence 338, Appl
9	46.6	2.8	516	10	US-09-960-352-5785
10	46.4	2.8	7047	13	US-10-240-453-259
11	46.4	2.8	17848	13	US-10-240-453-38
12	46.4	2.8	17848	15	US-10-239-676-28
13	45.2	2.7	616	13	US-10-027-632-233686
14	45.2	2.7	616	14	US-10-027-632-233686
15	45	2.7	502	13	US-09-814-353-17272

C 16	44.8	2.7	1271	10	US-09-938-842A-2829	Sequence 2829, Ap
C 17	44.8	2.7	1271	12	US-09-938-842A-2829	Sequence 2829, Ap
C 18	44.6	2.7	556	13	US-09-814-353-4676	Sequence 4676, Ap
C 19	44.6	2.7	556	13	US-09-814-353-10975	Sequence 10975, A
C 20	44.6	2.7	2000	10	US-09-938-842A-4038	Sequence 4038, Ap
C 21	44.6	2.7	2000	12	US-09-938-842A-4038	Sequence 4038, Ap
C 22	44.2	2.6	19734	13	US-10-311-455-1906	Sequence 1906, Ap
C 23	44.2	2.6	113515	13	US-10-311-455-2148	Sequence 2148, Ap
C 24	44	2.6	7498	13	US-10-311-455-230	Sequence 230, Appl
C 25	43.6	2.6	11473	13	US-10-311-455-1328	Sequence 1328, Ap
C 26	43.4	2.6	6533	13	US-10-240-453-257	Sequence 257, Appl
C 27	43.4	2.6	640681	10	US-09-790-988-1	Sequence 1, Appl
C 28	43.2	2.6	2938	15	US-10-198-846-9862	Sequence 9862, Ap
C 29	43.2	2.6	6175	13	US-10-311-455-1280	Sequence 1280, Ap
C 30	43	2.6	6244	13	US-10-311-455-458	Sequence 458, Appl
C 31	43	2.6	3673778	13	US-10-312-841-1	Sequence 1, Appl
C 32	42.8	2.6	5682	13	US-10-311-455-545	Sequence 545, Appl
C 33	42.8	2.6	5682	13	US-10-240-485-53	Sequence 53, Appl
C 34	42.8	2.6	9180	13	US-10-311-455-1938	Sequence 1938, Ap
C 35	42.8	2.6	3673778	13	US-10-312-841-2	Sequence 2, Appl
C 36	42.6	2.5	338	13	US-10-125-968-1452	Sequence 1252, Ap
C 37	42.6	2.5	730	15	US-10-198-846-2090	Sequence 2090, Ap
C 38	42.6	2.5	17137	13	US-10-311-455-164	Sequence 164, Appl
C 39	42.2	2.5	6668	13	US-10-311-455-1192	Sequence 1192, Ap
C 40	42.2	2.5	7312	13	US-10-311-455-1788	Sequence 1788, Ap
C 41	42.2	2.5	8693	13	US-10-311-455-1483	Sequence 1483, Ap
C 42	42.2	2.5	8693	15	US-10-172-086-37	Sequence 37, Appl
C 43	42.2	2.5	10138	13	US-10-240-453-218	Sequence 218, Appl
C 44	42.2	2.5	73308	10	US-09-954-456-2276	Sequence 2276, Ap
C 45	42	2.5	1960	10	US-09-938-842A-4687	Sequence 4687, Ap

ALIGNMENTS

RESULT 1

US-09-938-842A-3729
Sequence 3729, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 3729
LENGTH: 1677
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-3729

Query Match 100.0%; Score 1677; DB 10; Length 1677;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGTAAAGCGTTTACTATGTTTATATGCGACGGAAGATATGCGATGTTGGATCG 60

Db 1 GGTAAAGCGTTTACTATGTTTATATGCGACGGAAGATATGCGATGTTGGATCG 60

Qy 61 TTTTTCAGATCATCAAGGCTCCTACAGATTTCTTACGAGGATGTTTCAGGCTTTGTTA 120

Db 61 TTTTTCAGATCATCAAGGCTCCTACAGATTTCTTACGAGGATGTTTCAGGCTTTGTTA 120

b 121 GAAATTTGTTTATTTGCAACAGGTAGAGAACATTAACCATAGACAGATGTATCTGAAGAGA 180
y 181 TAAGCTTCTCTATGCTTAAGAAATGGACCGATACAGTAATAAACAAGCATCATTAAGAT 240
b 181 TAAGCTTCTCTATGCTTAAGAAATGGACCGATACAGTAATAAACAAGCATCATTAAGAT 240
y 241 TAAATGGTTTGAAGAAATACATACATTTATTTATGTGAAATTTGTGTGTAGTGAAGAT 300
b 241 TAAATGGTTTGTGAAGAAATACATACATTTATTTATGTGAAATTTGTGTGTAGTGAAGAT 300
y 301 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTTATGATGCTGGCG 360
b 301 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTTATGATGCTGGCG 360
y 361 TAAATGAATGGTATGCTGATGTGTAGGCAAAAGTTGTGTGCTCGCAAAATTTACCAATTC 420
b 361 TAAATGAATGGTATGCTGATGTGTAGGCAAAAGTTGTGTGCTCGCAAAATTTACCAATTC 420
y 421 CACTGTGTGAGACCGCAATCTGTAATCGGAAAGGTGGAACCCACTTGGTTTAACTTTT 480
b 421 CACTGTGTGAGACCGCAATCTGTAATCGGAAAGGTGGAACCCACTTGGTTTAACTTTT 480
y 481 AAGCTTAAAGGTTACTACCGGTTTGAACCGTTTATATTTTGGTGTATTAATCTTAATCCC 540
b 481 AAGCTTAAAGGTTACTACCGGTTTGAACCGTTTATATTTTGGTGTATTAATCTTAATCCC 540
y 541 GGATCCGTTTGTGTTAATCTCAAGGCCAGTTATGCCCAATATTTTGAATTTTGTAGTG 600
b 541 GGATCCGTTTGTGTTAATCTCAAGGCCAGTTATGCCCAATATTTTGAATTTTGTAGTG 600
y 601 GGTAGGAATGCTGGGGTGCATAGTTGGGCTAGCCCTCAACATGTGTGGAATCGAG 660
b 601 GGTAGGAATGCTGGGGTGCATAGTTGGGCTAGCCCTCAACATGTGTGGAATCGAG 660
y 661 AGAGTAGGCTCAGCTCAGGCCCAATTCACATTTTCTTTTGTAGCTTTCTTTTCTGG 720
b 661 AGAGTAGGCTCAGCTCAGGCCCAATTCACATTTTCTTTTGTAGCTTTCTTTTCTGG 720
y 721 TGCTTACGCTCCTTTCTGTGTGCTGATGTATGTAAGTAGTATGTAAGTGTGTTTCTGG 780
b 721 TGCTTACGCTCCTTTCTGTGTGCTGATGTATGTAAGTAGTATGTAAGTGTGTTTCTGG 780
y 781 CCGGAAACAGTACCAACGAATCAAAATAGTTTGAATCGGTACATCTAGTTACCGTGC 840
b 781 CCGGAAACAGTACCAACGAATCAAAATAGTTTGAATCGGTACATCTAGTTACCGTGC 840
y 841 AACTTACATCAATTTTCAATTTGATTTGATTTCTGATTTCTAGTTTCTGATTTTAAAT 900
b 841 AACTTACATCAATTTTCAATTTGATTTCTGATTTCTAGTTTCTGATTTTAAAT 900
y 901 CCGGATTTGACAGTACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 960
b 901 CCGGATTTGACAGTACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 960
y 961 AAAGGACGCAATAGGAGGATTTTGGAAATCTGGAAGAGAGATTTTCCATAGACACTA 1020
b 961 AAAGGACGCAATAGGAGGATTTTGGAAATCTGGAAGAGAGATTTTCCATAGACACTA 1020
y 1021 ATTAGTTTGGTGGCGAGCTTTGTGACCTACATTAATGGGTTCCACCCAGATG 1080
b 1021 ATTAGTTTGGTGGCGAGCTTTGTGACCTACATTAATGGGTTCCACCCAGATG 1080
y 1081 GCGTTACAGCTTTTCCATAAATTTAAAGTAAATCTTTTGGCTAACCAATTAAT 1140
b 1081 GCGTTACAGCTTTTCCATAAATTTAAAGTAAATCTTTTGGCTAACCAATTAAT 1140
y 1141 ATTGAATTTTCCAAACCATAGAAAGTTTAAATTTGATCGGATGAAATTTTGTAC 1200
b 1141 ATTGAATTTTCCAAACCATAGAAAGTTTAAATTTGATCGGATGAAATTTTGTAC 1200
y 1201 AAAGCTAGTATTTCAATTTGGAGGTGCTAGTAACTAGTAACTAGTAACTAGTAACTAGT 1260

Db 1201 AAAGCTAGTATTTCAATTTGGAGGTGCTAGTAACTAGTAACTAGTAACTAGTAACTAGT 1260
Qy 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTAACAAT 1320
Db 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTAACAAT 1320
Qy 1321 AAAGMAAACAATTTTGTGAAGAGAGAAATTAAGTTTACTGGAACCCATTTGACAGATG 1380
Db 1321 AAAGMAAACAATTTTGTGAAGAGAGAAATTAAGTTTACTGGAACCCATTTGACAGATG 1380
Qy 1381 TCCATATAATACTAGTAGAGAGATAGCAATGAAGTGAATTTGTTCAAGTGTACAA 1440
Db 1381 TCCATATAATACTAGTAGAGAGATAGCAATGAAGTGAATTTGTTCAAGTGTACAA 1440
Qy 1441 TCGAATGTTTCTTAAAGCTCATCGAACACATCAGGACCGTTGATTTTCCCGCATCA 1500
Db 1441 TCGAATGTTTCTTAAAGCTCATCGAACACATCAGGACCGTTGATTTTCCCGCATCA 1500
Qy 1501 AAAGGTTCAATPACTTCTCACTTGTGTTTCTCTCTATATATATATCTCGACGATC 1560
Db 1501 AAAGGTTCAATPACTTCTCACTTGTGTTTCTCTCTATATATATATCTCGACGATC 1560
Qy 1561 ACATTTAGTAATCTCTCTGAGCTGTAAACGGGTTAAACGATTTCTTTCCATTTGATCC 1620
Db 1561 ACATTTAGTAATCTCTCTGAGCTGTAAACGGGTTAAACGATTTCTTTCCATTTGATCC 1620
Qy 1621 GCTTTTAAACATCTCTGCTGCTCATCTCCACCGTTCGTTTCTCTCAGTATATTTTA 1677
Db 1621 GCTTTTAAACATCTCTGCTGCTCATCTCCACCGTTCGTTTCTCTCAGTATATTTTA 1677

RESULT 3

US-09-887-576-73/c

; Sequence 73, Application US/09887576

; Patent No. US20020144047A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, P.

; APPLICANT: Brown, D.

; APPLICANT: Chang, H.

; APPLICANT: Zhu, T.

; APPLICANT: Han, B.

; APPLICANT: Wang, X.

; APPLICANT: Cooper, Bret

; TITLE OF INVENTION: Promoters for regulation of plant expression

; FILE REFERENCE: 1360.001US1

; CURRENT APPLICATION NUMBER: US/09/887,576

; CURRENT FILING DATE: 2001-06-25

; PRIOR APPLICATION NUMBER: US 60/213,848

; PRIOR FILING DATE: 2000-06-23

; PRIOR APPLICATION NUMBER: US 60/214,087

; PRIOR FILING DATE: 2000-06-23

; PRIOR APPLICATION NUMBER: US 60/258,692

; PRIOR FILING DATE: 2000-12-29

; NUMBER OF SEQ ID NOS: 875

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 73

; LENGTH: 2000

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-887-576-73

Query Match 3.1%; Score 52.4; DB 10; Length 2000;
Best Local Similarity 47.1%; Pred. No. 0.067;
Matches 161; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

Qy 1075 AGTATGGCTTACAGCTTTTCCATTAATTAAGTAAATCTTTTTCCTAACCAATA 1134
Db 1019 ATTAGGTTCTGATCTTTTAACTATAAAAAAATACTATATGATGATGATGATG 960
Qy 1135 AATAATTTTGAATAATCTTTCCACCATAGAAAGTAAATTTGATCAGCGATGGAAT 1194
Db 959 AGAATTTGAAGTACGAACAGGCTCTTAAACCTGTTCTCTATTTTGGTATGATCAT 900

1195 TTGTACAAAGCTAGGTATTTTCATTTGGAGTGTACTAGTAAGTACTAAACCGA 1254
899 TTCAAACTTGAGATATTTAGCATTAACCTTACAGCAAAAGAAAAAAGCTTGCATTAC 840
1255 ATGAGTTTCTGATTTGGATTTTGAAGCTTTTCTTCTTGGTTTAAAAACAAGTATATTACTA 1314
839 AAGTGTGTTTTTTTTTTGTTTTTGAACACAGTTTAAAGAACCGAACTTGATCGATTATTATTA 780
1315 AACCAATAAAGAAAAACATTTTGTGAAAGAGAGAAATAAAGTTTCTGCGACCCCATTTGTAC 1374
779 GACAGCGAGCTACCCCTTTGTGTTGTTGACGATCCTATATTAAGTTTTCACGGAAC 720
1375 AGATGTCCTCATATATATCTATAGTAGAGATAGAGCAATGGA 1416
719 GCCTGACGGTTAATCGCATCGGTAGAGAAAGATCCGTCGA 678

RESULT 4

US-10-198-846-7035/c
; Sequence 7035, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: NRI-049
; CURRENT APPLICATION NUMBER: US/10/198.846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7035
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137, 182, 183, 184, 185, 186, 187, 188, 209, 211, 212, 217,
; LOCATION: 219, 224, 225, 236, 237, 238, 240, 241, 242, 248, 250, 252,
; LOCATION: 253, 259, 276, 279, 281, 284, 298, 304, 306, 315, 318, 319,
; LOCATION: 320, 321, 327, 338, 349, 350, 352, 355, 363, 378, 379
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 380, 386, 389, 393, 398, 399, 400, 408, 410, 414, 434, 441,
; LOCATION: 443, 446, 451, 459, 466, 488, 490, 497, 498, 499, 501, 503,
; LOCATION: 506, 509, 513, 517, 526, 527, 528, 529, 536, 550, 557, 562,
; LOCATION: 564, 565, 573, 576, 588, 599, 604, 607, 616, 617, 619
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 627, 628, 643, 649, 661, 668, 672, 680, 689, 706, 719,
; LOCATION: 722, 735, 739, 750, 763, 765, 769, 771, 772, 774, 784, 787,
; LOCATION: 790, 793, 792, 795, 798, 805, 814, 816, 819, 820, 822, 830,
; LOCATION: 832, 833, 838, 842, 847, 849, 850, 853, 856, 857
; OTHER INFORMATION: n = A,T,C or G

US-10-198-846-7035

Query Match 3.0%; Score 49.6; DB 15; Length 858;
Best Local Similarity 42.1%; Pred No. 0.2;
Matches 118; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

1088 AGCTTTTCCATAAATAAGTAAATCTTTTGGCTAACCAATAAATAATTTCGAA 1147
382 ATNNNTTGAAGAAAAATATAAAAAAANTNGNGGAATTAATAAATAATTTTTTTNAAA 323
1148 ATCTTCCACCATAGAAAAGTTAATTTGTATCGCATCGGATGGAATTTTGTACAAAGCTA 1307

322 GNNNCANAAATTAANNTNTTTTNGTNTAAANAAAAAANANNTNTTTTAATTTTT 263
1208 GGTATTTTCATTTGGAGTGTACTAGTAAGTACTAAACCAAGATGAGTTTCTGAT 1267
262 AAGNTTTTANNANANATTTNNNANNTTAAATTTTNTTAAANATTAANNCNTTTTT 203
1268 TTTGATTTTGAAGCTTTTCTTGGTTTAAAAACAAGTATATTACTTAACAAATAAAAAGAA 1327
202 TTTTNTTTTATATANNNNNNAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 143
1328 AACATTTTGTGAAAGAGAGAAATAAAGTTTCTGAGACCCC 1367
142 AAAANNAAGAAAAAATAAAAAAATAAAAAAATAAAAAAAGTACCTC 103

RESULT 5

US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:

; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312.841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
; US-10-312-841-2

Query Match 2.9%; Score 49.2; DB 13; Length 3673778;
Best Local Similarity 49.2%; Pred No. 19;
Matches 129; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

1096 CCATTAATTAAGTAATCTTTTTCCTTACCAATAAATAATTTGAAATCTTCC 1155
2005659 CCATAAATCTTAAAAAACCCTTAATCTTAAACCCCAAAAAAATTTCCAAAACTATA 2005600
1156 AACCATGAAAGTTAAATTTGATCAGCATGGAATTTTGTACAAAGCTAGGTATTC 1215
2005599 GCACAAACCCAAACAAACAAACATAAAAAACAAATATTAACCACTAACACG 2005540
1216 ATTTGGGAGTGTACTAGTAACTAGTAACTAACCAAGAGTTCCTGATTTGGATT 1275
2005539 ATATATAAAACATACCTTAATACATAATAATAACAAATTAACCTTTAATAAAT 2005480
1276 TTGAAGCTTTCTTAGGTTAAAAAACAAGTATATTACTTAACCAATAAAGAAAAACATT 1335
2005479 TCATAAATATATATATATATAATATAATTAATACATTAATAAATTAATAAATAAT 2005420
1336 TGTGAAAAGAGAAATAAAGTTT 1357
2005419 ACCAAAAATTTAAAAAAATTT 2005398

RESULT 6

US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:

; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312.841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2

Query Match 2.9%; Score 47.8; DB 10; Length 2958;
Best Local Similarity 50.7%; Pred. No. 1;
Matches 115; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
1228 ACTAGTAAGTAACTAGTACTACCAAGATGAGTTCTGATTTGGATTTTGAAGCTTTTC 1287
; Patent NO. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND


```

QY      1208  GGTATTTTCATTTCGGAGTGTACTAGTACTAGTACTAACCAGAAATGAGTTTCTCAT 1267
DB      9927  ATAAAAATAAATACTAAAATTAACCAAAAAAAAAATAAATAATTTAACCAATTTTTAATTTTCTTTT 9868
QY      1266  TTTCGAATTTTGAGCTTTTCTTAGGTGTTAAAAACAAGTATATTTACTAAACAATAAAAAAGAA 1327
DB      9867  TATAAAAAAATTAACACATAATTTTAAATTTAACCAAAATAATCTAATAAAAAAATAAAAA 9808
QY      1328  AAAACATTTTGTGAAAAGAGAAATAAAGTTTACTCGACCCCAATTGTACAGATGGTCCCATA 1387
DB      9807  AATTAAATAAATTCGAAACAAAAATAAAATTTACTAAAATAATTTTTTAAACAAAAAATAAATA 9748
QY      1388  ATAATCTCTGATAGAAGATGAGCAATCGGAAAGTGATTTGTTTCACTGGTGACAAATCGGAAT 1447
DB      9747  AAATCTAATTAACCAATAAAAAATAATTTAATCTTAATAAAAAACAAATTCATCCTAAT 9688

RESULT 13
US-10-027-632-233686
; Sequence 233686, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ FOR WINDOWS VERSION 4.0
; SEQ ID NO 233686
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-233686

Query Match          2.7%; Score 45.2; DB 13; Length 616;
Best Local Similarity 62.3%; Pred. No. 1.8;
Matches 71; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY      1325  GAAAAACATTTTGTGAAAGAGAAATAAAGTTTCTTGGACCCCAATTGTACAGATGGTCCC 1384
DB      476  GAGACTCTTTGTGCAAAAAAATAAAGTTTATTGGACCCCAATTGTAGATACAGTGCC 535
QY      1385  ATAATACTCTGATAGAGATAGACAAATCGGAAAGTGATTTTTCACGTGGTACAAATCGGAAT 1438
DB      536  CTGAGATCAAGGATAGGATACAAATAGGATAGGCCTTTGCCCTCAATGGAC 589

RESULT 14
US-10-027-632-233686
; Sequence 233686, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30

```

```

; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 233686
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-233686

Query Match          2.7%; Score 45.2; DB 14; Length 616;
Best Local Similarity 62.3%; Pred. No.1.8;
Matches 71; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1325 GAAAAACATTTTGAAAAGAGAAATAAAGTTTCTACGGACCCCATCTGTACAGATGGTCCC 1384
Db 476 GAGACTCTTTTGTCAAAAAATAAAGTTTATTGGACCCCATCTGTACATACAGTGC 535

QY 1385 ATAATACTGTAGAGATAGAGCAATGAAAGTGTATTTTTCACGTGGTAC 1438
Db 536 CTGAGATGAAGATAGGATCAACAACAAATAGCATAGGCCTTTTCGCTCAATGCAC 589

```

```

RESULT 15
US-09-814-353-17272/c
; Sequence 17272, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814.353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17272
; LENGTH: 502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 86
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-17272

```

	Query Match	2.7%	Score 45;	DB 13;	Length 502;	
	Best Local Similarity	49.4%;	Pred. No. 1.8;			
	Matches 117;	Conservative 0;	Mismatches 120;	Indels 0;	Gaps 0;	
QY	1119	TTTGGCTTAACCAATAAAATTAATTCGAAATCTTCCAACCATAGAAAAGTTAAATTGA	1178			
DB	347	TTTGGTTAAATTTATCTTTTTTATATATTTTCCGCCCCGCCCATTTTTTTTTTA	288			
QY	1179	TCAGGATGGAAATTTTTGTGACAACTAGTAGTATTTCATTGGGAGTGTACTAGTAACTA	1238			
DB	287	TTCCCTTTTTTAAATTTTGCCTCCCAACAAAATTTTTTTTTTTTACATTTTAAAAA	228			
QY	1239	GTAAGTACTACCAAGAATGAGTTTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAA	1298			
DB	227	TGAAAAATAAAAAAATTTTTTTTTTAAAAAATAAAAAAATAAAAAAATAAAAA	168			
QY	1299	AACAGTATATTACTAAACAATAAAAAAGAAAAACATTTGTGAAAAAGAGAAATAAAGT	1355			
DB	167	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAGT	111			

Search completed: February 2, 2004, 01:10:49
Job time : 656.401 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: February 1, 2004, 21:36:50 ; Search time 4063.41 Seconds
(without alignments)
10782.626 Million cell updates/sec

title: US-09-938-842A-1034

effect score: 1071

sequence: 1 atggcgacaattcgagct.....caoggtcatcgaaacctga 1071

coring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

searched: 2888711 seqs, 20454813386 residues

total number of hits satisfying chosen parameters: 5777422

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

GenBank:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.mu.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rtd.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1071	100.0	1071	6	AX506339	Sequence
2	1071	100.0	1071	8	AF370606	Arabidops
3	1071	100.0	1102	8	AY113981	Arabidops
4	1071	100.0	1650	8	AY056214	Arabidops
5	1071	100.0	91854	8	AC003580	Arabidops
C	157.8	14.7	882	8	BT008779	Arabidops
6	157.8	14.7	1046	8	AK118579	Arabidops
7	157.8	14.7	1156	8	AY086379	Arabidops
8	157.8	14.7	1156	8	AY086379	Arabidops
9	157.8	14.7	16203	8	AB025623	Arabidops
10	129	12.0	1470	8	AY103304	Arabidops
11	129	12.0	1755	8	AY058874	Arabidops
12	129	12.0	101154	8	ATFLP2	Arabidops
13	124.6	11.6	451	11	EX322174	Arabidops
14	124.6	11.6	978	8	AY081706	Arabidops
15	124.6	11.6	1440	8	AY056384	Arabidops
16	124.6	11.6	87148	8	AC013289	Arabidops
17	124.6	11.6	1432	8	AF165924	Gossypium
18	121.6	11.4	1206	8	BT008493	Arabidops
19	121.6	11.4	1479	8	AY081344	Arabidops
20	121.6	11.4	85961	8	AC079131	Arabidops
C	121.6	11.4	103637	8	AC079604	Arabidops
21	118	11.0	976	8	BT006043	Arabidops
22	118	11.0	1081	6	AX027400	Sequence
23	118	11.0	1085	8	ATH291749	Arabidops
24	118	11.0	1204	8	AK118178	Arabidops
25	118	11.0	73866	8	AB026649	Arabidops
C	117.8	11.0	1520	8	DB7261	Oryza sativ
27	117.8	11.0	154255	2	AP004632	Oryza sat
28	117.8	11.0	105437	2	CNS07YPE	Oryza sat
C	113.8	10.6	139487	2	AP003513	Oryza sat
30	113	10.6	164978	2	AP003569	Oryza sat
31	113	10.6	164978	2	AP003569	Oryza sat
32	111.2	10.4	720	8	AY072820	Arabidops
33	111.2	10.4	941	8	AF412074	Arabidops
34	111.2	10.4	125411	8	ATF8L15	Arabidops
C	110.4	10.3	74508	2	AP006175	Oryza sat
35	110.4	10.3	155102	2	AP005848	Oryza sat
C	108.4	10.1	95845	8	AF411807	Lycopersi
37	108.4	9.8	1549	8	AB071804	Oryza sat
38	105	9.8	193388	2	AC120539	Oryza sat
39	105	9.8	138882	8	AP004223	Oryza sat
40	104.6	9.8	146220	8	AP004672	Oryza sat
41	104.6	9.6	1159	8	AY150488	Arabidops
42	102.8	9.6	1508	8	AY080776	Arabidops
-43	102.8	9.6	38529	8	AC069273	Arabidops
C	102.8	9.3	1304	8	AK118679	Arabidops

ALIGNMENTS

RESULT 1
AX506339
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AX506339 1071 bp DNA linear PAT 27-SEP-2002
Sequence 1034 from Patent WO0216655.

AX506339

AX506339.1 GI:23387576

Arabidopsis thaliana (thale cress)
Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsals.

Harper, J.F., Kreps, J., Wang, X. and Zhu, T.

Stress-regulated genes of plants, transgenic plants containing

same, and methods of use
Patent: WO 021655-A 1034 28-FEB-2002;
The Scripps Research Institute (US) ; Syngenta Participations AG
(CH)

JOURNAL

FEATURES
source Location/Qualifiers
1. 1071
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"

BASE COUNT 287 a 256 g 240 t

ORIGIN

Query Match 100.0%; Score 1071; DB 6; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCAATTCAGAGCTTGAGAGTTGCGAGGCAAGATCAAACTCTAGAGCGTT 60
DB 1 ATGGCGCAATTCAGAGCTTGAGAGTTGCGAGGCAAGATCAAACTCTAGAGCGTT 60

QY 61 GATCTAACCATCATCAACGGCGTCAGAAACGTCGAAACCTTCAAGAGTAAAT 120
DB 61 GATCTAACCATCATCAACGGCGTCAGAAACGTCGAAACCTTCCBAGTAAT 120

QY 121 CCACAGTGTCTCGAGCCCAAGGCGGAGCCGGTGATGCGTCTGTTCAATGCTTTA 180
DB 121 CCACAGTGTCTCGAGCCCAAGGCGGAGCCGGTGATGCGTCTGTTCAATGCTTTA 180

QY 181 GCTCCACCGCTTCGACAGGACCAACCATCAAGAGAGCTTCGACTAAAGACGTCACAG 240
DB 181 GCTCCACCGCTTCGACAGGACCAACCATCAAGAGAGCTTCGACTAAAGACGTCACAG 240

QY 241 AAGTTGAAGAGAGGAGGAAGATACGATGCTGCCACGTTGCGGCTAGATTTT 300
DB 241 AAGTTGAAGAGAGGAGGAAGATACGATGCTGCCACGTTGCGGCTAGATTTT 300

QY 301 CAATTAACGAGTTAGTCTCAATTCGAGCGGCAAGATTCGTTGTTGGAG 360
DB 301 CAATTAACGAGTTAGTCTCAATTCGAGCGGCAAGATTCGTTGTTGGAG 360

QY 361 AAGCTGAGCGCGGATTTAGCGCCGACCGGTCACGGGACCGGTTCCCGCCATCGCATG 420
DB 361 AAGCTGAGCGCGGATTTAGCGCCGACCGGTCACGGGACCGGTTCCCGCCATCGCATG 420

QY 421 TCGGTTAACGGAACCTTAAATTCGACGACGACGACGCTGATTCGATGCGTGAA 480
DB 421 TCGGTTAACGGAACCTTAAATTCGACGACGACGACGCTGATTCGATGCGTGAA 480

QY 481 AATCTGATGAAGAAACGTTAAACGACCTTCTAAACAGTGTATATAGACATAAGCGAC 540
DB 481 AATCTGATGAAGAAACGTTAAACGACCTTCTAAACAGTGTATATAGACATAAGCGAC 540

QY 541 GCCGTTTTCAGTCTCCGTTTGTAGTCCAAATTCGACGACGACGACGATCCAACTCG 600
DB 541 GCCGTTTTCAGTCTCCGTTTGTAGTCCAAATTCGACGACGACGACGATCCAACTCG 600

QY 601 CAAGCTCTGGCATCATCACTGTGGCTCAGCAACTCTTCGCGGCAAGATGTATCGATG 660
DB 601 CAAGCTCTGGCATCATCACTGTGGCTCAGCAACTCTTCGCGGCAAGATGTATCGATG 660

QY 661 TGGGCTATTCCATCAACGCAATGATTCGACGCTCGAGCTTCTTCTGTATCCACAA 720
DB 661 TGGGCTATTCCATCAACGCAATGATTCGACGCTCGAGCTTCTTCTGTATCCACAA 720

QY 721 ATCGCTGTCGTCGATCAGCTCAGTATTAGCTTTTCCGCGCGCGCTGTCGCG 780
DB 721 ATCGCTGTCGTCGATCAGCTCAGTATTAGCTTTTCCGCGCGCGCTGTCGCG 780

QY 781 TCGTCTTACGTCGCGCTGTTTCAACAGGCTTCCACGATGCTAGACACCTCTTACAA 840
DB 781 TCGTCTTACGTCGCGCTGTTTCAACAGGCTTCCACGATGCTAGACACCTCTTACAA 840

QY 841 GTTGTTCACAGCAGCGGCTTGTATCGCTTTCAGACGTTAGCGGTCGAATTTATCAAG 900

841 GTTGTTCACAGCAGCGGCTTGTATCGCTTTCAGACGTTAGCGGTCGAATTTATCAAG 900

QY 901 CGCAGCTGCGGTTATGGCTCCGAGCTCAAGCTCAGGCGTCAACACCGGTAGTTCATGCTCA 960
DB 901 CGCAGCTGCGGTTATGGCTCCGAGCTCAAGCTCAGGCGTCAACACCGGTAGTTCATGCTCA 960

QY 961 ATTCAACACACACGAGCGCACAGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
DB 961 ATTCAACACACACGAGCGCACAGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020

QY 1021 GAGTTCCAGGTTTCATGAGCACCAACACGACGCTCATCGAACCACTGA 1071
DB 1021 GAGTTCCAGGTTTCATGAGCACCAACACGACGCTCATCGAACCACTGA 1071

RESULT 2
AF370606 1071 bp mRNA linear PLN 30-APR-2001
LOCUS Arabidopsis thaliana putative PCF2-like DNA binding protein
DEFINITION (At2g45680; F17K2.21) mRNA, complete cds.
ACCESSION AF370606
VERSION AF370606.1 GI:13877694
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 1071)
AUTHORS Lan, B., Southwick, A., Karlin-Neumann, G., Nguyen, M., Miranda, M.,
Pam, C.J., Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R.,
Chung, M.K., Kim, C., Lin, J., Liu, X., Pham, P.K., Sakano, H.,
Shinn, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.
Direct Submission
Submitted (17-APR-2001) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
e-mail for correspondence: arabesequence.stanford.edu
This clone was isolated by RT-PCR.
FEATURES
source Location/Qualifiers
1. 1071
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="II"
/clone="CS15"
/notes="ecotype: Columbia"
1. 1071
/gene="At2g45680; F17K2.21"
1. 1071
/gene="At2g45680; F17K2.21"
/codon_start=1
/product="putative PCF2-like DNA binding protein"
/protein_id="AAK43925.1"
/db_xref="GI:13877695"
/translation="MATIOKLEEVAGKDOTLRVLDLTIINGVNVETSPFOVNPVTS
LEPKAEVMPSPMSLAPSPGPIPKRASTKHRTKVEGRGRIRMPATCAARIFOL
TRELGKSDGETTIRMLERPAIIATGCTGTPALAMSVNGTLKPTTTNADSDMG
NLMKKGRFNSSEYIDISDAVSAGSLAPAIATTTIQQPQALASSTVAQQLPQGM
PMWAPSNAMIPVGFAPFLIPQIAGPSNQPLLAFAAASPSYSVAQVQASTWARP
PPLQVPSGFMFVSVDVSNLSNLRATSVAPSSSGVTTGSSSSSIATTTTILRDFSL
EIVEKQELHQMSTTTARSNH"
BASE COUNT 287 a 256 g 240 t
ORIGIN

Query Match 100.0%; Score 1071; DB 8; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCAATTCAGAGCTTGAGAGTTGCGAGGCAAGATCAAACTCTAGAGCGTT 60
DB 1 ATGGCGCAATTCAGAGCTTGAGAGTTGCGAGGCAAGATCAAACTCTAGAGCGTT 60

61 GATCTAACCATCATCAAGCGGCTCAGAAACGTCGAAACTTCAAGACCTTTCCAAAGTAAT 120
b 61 GATCTAACCATCATCAAGCGGCTCAGAAACGTCGAAACTTCAAGACCTTTCCAAAGTAAT 120
y 121 CCCACAGTAGTCTCAGCGCCCAAGCGGAGCGGCGTATGCGGTGTTTCAATGCTTTA 180
b 121 CCCACAGTAGTCTCAGCGCCCAAGCGGAGCGGCGTATGCGGTGTTTCAATGCTTTA 180
y 181 GCTCCACCGTCTTCACAGGACCCACATTTGAAGAGAGTTCGACTAAAGACCGTCACACG 240
b 181 GCTCCACCGTCTTCACAGGACCCACATTTGAAGAGAGTTCGACTAAAGACCGTCACACG 240
y 241 AAGGTGAAGGAGGAGGAGGAGGATCGGATCGCTGCGGCTAGGATTTT 300
b 241 AAGGTGAAGGAGGAGGAGGAGGATCGGATCGCTGCGGCTAGGATTTT 300
y 301 CAATTAACCTCAGAGTGTAGTTCACAAATCCGACGCGGAAACGATTCGGTGGTGTGGAG 360
b 301 CAATTAACCTCAGAGTGTAGTTCACAAATCCGACGCGGAAACGATTCGGTGGTGTGGAG 360
y 361 AACGCTGACCGCGGATTAATACCGCCACGCGTACGGAAACGGTTCGCCCATCCCATG 420
b 361 AACGCTGACCGCGGATTAATACCGCCACGCGTACGGAAACGGTTCGCCCATCCCATG 420
y 421 TCGGTTACGGAACCTTAAATCCGACGAGGAGGAGGATTCGATTCGATTCGGTGA 480
b 421 TCGGTTACGGAACCTTAAATCCGACGAGGAGGAGGATTCGATTCGATTCGGTGA 480
y 481 AATCTGATGAAGAAACGTAACGACCTTCTACAGTGAATATAGACATTAAGCGAC 540
b 481 AATCTGATGAAGAAACGTAACGACCTTCTACAGTGAATATAGACATTAAGCGAC 540
y 541 GCGGTTTACGTTTCCGTTTACTGCTCAATTCGACGAGCAACGATCCACCTCCG 600
b 541 GCGGTTTACGTTTCCGTTTACTGCTCAATTCGACGAGCAACGATCCACCTCCG 600
y 601 CAAGCTTGGCATATCACTGTGCTCAGCAACTTCTGCGCGAGGAATGTATCCGATG 660
b 601 CAAGCTTGGCATATCACTGTGCTCAGCAACTTCTGCGCGAGGAATGTATCCGATG 660
y 661 TGGGCTATTCATCAACGCAATGATTCGACGCTGCGAGCTTTCTTCTGATTCACAA 720
b 661 TGGGCTATTCATCAACGCAATGATTCGACGCTGCGAGCTTTCTTCTGATTCACAA 720
y 721 ATCGCTGCTCGCTCGATCAGCTCAGTATATAGCTTTTCCGCGCGCTGCTTCGCG 780
b 721 ATCGCTGCTCGCTCGATCAGCTCAGTATATAGCTTTTCCGCGCGCTGCTTCGCG 780
y 781 TCGTCTTACGTCGCGCTGTTCAACAGGCTTCCAGGCTGAGTACACCTCCTTTACAA 840
b 781 TCGTCTTACGTCGCGCTGTTCAACAGGCTTCCAGGCTGAGTACACCTCCTTTACAA 840
y 841 GTTGTTCCAGCAGCGCTTTGATTCGTTTCCAGAGTTCAGAGTTCGATTCAGGA 900
b 841 GTTGTTCCAGCAGCGCTTTGATTCGTTTCCAGAGTTCAGAGTTCGATTCAGGA 900
y 901 GCGAGTCTGCTTATGCTCCGAGCTCAAGCTCAGCGTAAACGCGTATTCATCGTCA 960
b 901 GCGAGTCTGCTTATGCTCCGAGCTCAAGCTCAGCGTAAACGCGTATTCATCGTCA 960
y 961 ATTGCAACCAACGAGCAGCAGCTGAGAGATTTCTCCCTAGAGATATCAGAAACAA 1020
b 961 ATTGCAACCAACGAGCAGCAGCTGAGAGATTTCTCCCTAGAGATATCAGAAACAA 1020
y 1021 GAGCTTCCAGCTTATGAGCAGCAGCAGCAGCGGTCATGAACCACTGA 1071
b 1021 GAGCTTCCAGCTTATGAGCAGCAGCAGCAGCGGTCATGAACCACTGA 1071

RESULT 3
AY113981
OCUS

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Arabidopsis thaliana putative PCF2 DNA binding protein (At2g45680)
mRNA, complete cds.
AY113981
Version 1
GI:21281084
FLI CDNA
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis

REFERENCE
AUTHORS

1 (bases 1 to 1102)
Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M.,
Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L.,
Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L.,
Carrinci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Open Reading Frame (ORF) Clones

TITLE
JOURNAL

Unpublished
2 (bases 1 to 1102)
Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M.,
Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L.,
Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L.,
Carrinci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission

TITLE
JOURNAL

Submitted (16-MAY-2002) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
The RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: RIKEN
Arabidopsis Full-Length cDNA): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

COMMENT

The Salk, Stanford, PEGC (SSP) Consortium members constructed and
sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada, K.,
Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M.,
Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C.,
Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Chen, H.,
Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J.,
Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PEGC) and Seki, M. (RIKEN GSC) contributed equally
to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP
/PEGC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.

FEATURES
source

1..1102
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="U10678"
/note="This clone is in pUNI 51."
ecotype: Columbia
1..1102
/gene="At2g45680"
1..1071
/gene="At2g45680"
/codon_start=1
/evidence=experimental
/product="putative PCF2 DNA binding protein"
/protein_id="AA045029.1"
/db_xref="GI:21281085"
/translation="WATIQLEEVAGKDTLRAVLTITNGVNRVETSRPFQVNFVTS

gene

1..1071
/gene="At2g45680"

CDS

1..1071
/gene="At2g45680"
/codon_start=1
/evidence=experimental
/product="putative PCF2 DNA binding protein"
/protein_id="AA045029.1"
/db_xref="GI:21281085"
/translation="WATIQLEEVAGKDTLRAVLTITNGVNRVETSRPFQVNFVTS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

source
1. 1660


```
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF07-08-P04 (RI0678)"
/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert.
ecotype: Columbia"
1..1660
/gene="At2g45680"
1..102
/gene="At2g45680"
103..1173
/gene="At2g45680"
/codon_start=1
/evidence=experimental
/product="putative PCF2 DNA binding protein"
/protein_id="AA07063.1"
/db_xref="GI:15810351"
/translation="WATIQLEEVAGKDOTLRVLDLTIINGVNRVETSRPQVNPVTS
LEPKAEVPWPSFMSLAPSSGTPELKAETKDSHTKVEGRIRMPATCAARIFOL
TRELGKSDGETIRWLENAPAIATGTGTVAIMSVNGTLIKIPITINADSDMGH
NMKKRKPNSEYIDISDAVSASSGLAPIATTTIOPOALASSTVAQQLPQGY
PMWALPSNMTPTVGAFLIPQIAGPSNPQOLAPFAAASPSYVAQQAQSTARP
PELVVPSGFSVSDVSGSLRSATSVAPSSSGVTGSSSIATTTTTLRDFSL
EIVEKELHOFMSITTRSSNH"
1174..1660
/gene="At2g45680"
464 a 400 c 385 g 411 t

3'UTR
ASE COUNT 464 a 400 c 385 g 411 t
RIGIN
Query Match 100.0%; Score 1071; DB 8; Length 1660;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 ATGGCGACATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCTAAGAGCGGTT 60
b 103 ATGGCGACATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCTAAGAGCGGTT 162
Y 61 GATCTAACCATCATCAACGGCTCAGAACGTCGAAACTTCAAGACCTTCCAGTAAT 120
b 163 GATCTAACCATCATCAACGGCTCAGAACGTCGAAACTTCAAGACCTTCCAGTAAT 222
Y 121 CCCAGTGTAGTCTCGAGCCCAAGCGGCGGTGATGCCGTCTGTTCAATGCTTTA 180
b 223 CCCAGTGTAGTCTCGAGCCCAAGCGGCGGTGATGCCGTCTGTTCAATGCTTTA 282
Y 181 GCTCCACCGTCTCGACAGGACCACTTGAAGAGCTTGCAGTAAAGCCGTCACAG 240
b 283 GCTCCACCGTCTCGACAGGACCACTTGAAGAGCTTGCAGTAAAGCCGTCACAG 342
Y 241 AAGGTTGAAGAGGAGGAGGATAGCGATCGCTGCGACGTGCGGCTAGGATTTT 300
b 343 AAGGTTGAAGAGGAGGAGGATAGCGATCGCTGCGACGTGCGGCTAGGATTTT 402
Y 301 CAATTAACTCGAGAGTGTGTCACAAATCCGACGGCGGAAACGATTCGGTGTGAG 360
b 403 CAATTAACTCGAGAGTGTGTCACAAATCCGACGGCGGAAACGATTCGGTGTGAG 462
Y 361 AAGCGTGAGCGCGGATATAGCGGACGGGTAGCGGACGGTTCCGCGCATCGCATG 420
b 463 AAGCGTGAGCGCGGATATAGCGGACGGGTAGCGGACGGTTCCGCGCATCGCATG 522
Y 421 TCGGTTAAGCGAACCTTAAATATCCCGACGACGACGACGATTCGTATATGGGTGAA 480
b 523 TCGGTTAAGCGAACCTTAAATATCCCGACGACGACGACGATTCGTATATGGGTGAA 582
Y 481 AATCTGATGAAGAGAGAGTAAACGACCTTCACTGATGATATAGACATAGCGAC 540
b 583 AATCTGATGAAGAGAGAGTAAACGACCTTCACTGATGATATAGACATAGCGAC 642
Y 541 GCGGTTTTCAGCTTCTCCGGTTTTCAGTCCCAATGTCAGGACGACGACGATCCACCTCCG 600
```

```
Db 643 GCGGTTTTCAGCTTCTCCGGTTTTCAGTCCCAATGTCAGACGACGATCCAACTCCG 702
Qy 601 CAAAGCTCTGCGCATCATCCACTGTGGCTCAGCAACTTCTCCGCGCAAGGATGATCCGATG 660
Db 703 CAAAGCTCTGCGCATCATCCACTGTGGCTCAGCAACTTCTCCGCGCAAGGATGATCCGATG 762
Qy 661 TGGGCTATTTCATCAACCGCAATGATTCGAGCGGTCCGAGCTTCTTCTGATTCACAA 720
Db 763 TGGGCTATTTCATCAACCGCAATGATTCGAGCGGTCCGAGCTTCTTCTGATTCACAA 822
Qy 721 ATGCTGTGTCGCTCGAATCAGCTCAGTTTATGCTTTTCCGCGCGCGCTGCTTCGCG 780
Db 823 ATGCTGTGTCGCTCGAATCAGCTCAGTTTATGCTTTTCCGCGCGCGCTGCTTCGCG 882
Qy 781 TCGTCTTACGTGCGCGCTGTTCAACAGGCTTCAGATGGCTAGACCACTCTTTACAA 840
Db 883 TCGTCTTACGTGCGCGCTGTTCAACAGGCTTCACAGATGGCTAGACCACTCTTTACAA 942
Qy 841 GTTGTTCCTCAAGCAGCGGCTTGTATCCGTTTCAGACGTTTTCAGGTTTCAATTTACA 900
Db 943 GTTGTTCCTCAAGCAGCGGCTTGTATCCGTTTCAGACGTTTTCAGGTTTCAATTTACA 1002
Qy 901 GCGACGTGCTTATGCTCCGAGCTCAAGCTCAGCGTAAACCGGTAGTTTCATCGTCA 960
Db 1003 GCGACGTGCTTATGCTCCGAGCTCAAGCTCAGCGTAAACCGGTAGTTTCATCGTCA 1062
Qy 961 ATTGCAACCAACGACGCGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
Db 1063 ATTGCAACCAACGACGCGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1122
Qy 1021 GAGCTTCCAGCTTATGAGACCAACAGCAGCGTTCATGCAACCACTGA 1071
Db 1123 GAGCTTCCAGCTTATGAGACCAACAGCAGCGTTCATGCAACCACTGA 1173
```

RESULT 5

```
AC003680/c
LOCUS
DEFINITION
  Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence,
  complete sequence.
ACCESSION
  AC003680
VERSION
  AC003680.3 GI:20197048
KEYWORDS
  HTG.
ORGANISM
  Arabidopsis thaliana (thale cress)
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
  1 (bases 1 to 91854)
  Wu, D., Maiti, R., Renning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
  Earnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
  Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence
  Unpublished
  2 (bases 1 to 91854)
  Lin, X.
  Direct Submission
  Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
  Medical Center Dr., Rockville, MD 20850, USA
  3 (bases 1 to 91854)
  Town, C.D. and Kaul, S.
  Direct Submission
  Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
  Medical Center Dr., Rockville, MD 20850, USA, cdtown@igr.org
  On Apr 18, 2002 this sequence version replaced GI:6598396.
  Address all correspondence to: atetigr.org
```

BAC clone F17K2 is from Arabidopsis thaliana chromosome 2 and is near the molecular marker (S) FLS.
The orientation of the sequence is from SP6 to T7 end of the BAC clone.
Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge).

<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkEM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pettea, <http://www.tigr.org/softlab/glimmer.htm>/glimmerM.html, and GeneSplicer (Mihaela Pettea and Steven Salzberg, contact npettea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/cdb/tgr.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST or EST similarity are named as unknown proteins. Genes without protein or EST similarity that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

```

1. 91854
   /organism="Arabidopsis thaliana"
   /mol_type="genomic DNA"
   /cultivar="Columbia"
   /db_xref="taxon:3702"
   /chromosome="2"
   /map="C1C02B07"
   /clone="F17K2"
   /complement(1..1281)
   /note="overlap with BAC clone F4L23 (AC002387:1..1281)."
57..747
   /gene="At2g45480"
   /note="synonym: F17K2.27"
   join(<57..242,335..747)
   /gene="At2g45480"
   join(57..242,335..747)
   /gene="At2g45480"
   /note="Protein sequence is in conflict with the conceptual translation"
   /codon_start=1
   /product="hypothetical protein"
   /protein_id="AAF18607.2"
   /db_xref="GI:20197053"
   /translation="DESKPVSQKFFPEVSDKLEAKFSNNKNDIIARSREKMNVM
NGGLFHGFSPDTVQERQVETDNEPFCRGRTDGKWKRCSDVLSGGKYCD
KIMHRGMKKHPDVTNSENAGFSPLVETAVRSVWFCDDGDKHGSVWVGITLPR
VDEKSTSSCSSTDTTITDIALRGEDDDEYLSLSPGV"
   /complement(1002..2388)
   /gene="At2g45480"
   /note="synonym: F17K2.2; contains a protein kinase domain profile (PDOC00100)"
   /complement(join(<1002..1112,1208..1280,1365..1580,
1655..1750,1886..2128,2260..22388))
   /gene="At2g45480"
   /complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..2388))
   /gene="At2g45480"
   /codon_start=1
   /product="putative protein kinase"
   /protein_id="AAC06151.1"
   /db_xref="GI:2979542"
   /translation="MSKSTSDAGNTKQMSLADPEIGRLPKGKFGVYLAREAKS
KIYALKVIFKQIEKYKHQRMWEICTSLRHPNIRLFGWHDNERIFLILEYA
HGELYGVKQNGHLEQQAAITYALSQALAYCHGKVIRHDIKPENLLIHGGRLLK
IADFGVSQSKRKTTCGDTLDYLAPEVMVENDHDYADVNTWLTGLICYEFYLYGPPPE
ARSQKDTFRILKIDLSLFFLTNVSEAKNLISQLLVKDPKRLSIEKIMQHPVIVKN
ADPKGVCASDI"
   /complement(3414..5980)
   /gene="At2g45500"
   /note="synonym: F17K2.3; predicted by genescan"
   /complement(join(<3414..3616,3639..3825,3916..3965,
4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,
5506..5585,5670..5980))
   /gene="At2g45500"

```

```

CDS
complement(join(3414..3616,3639..3825,3916..3965,
4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,
5506..5585,5670..5980))
   /gene="At2g45500"
   /codon_start=1
   /product="hypothetical protein"
   /protein_id="AAC06152.1"
   /db_xref="GI:2979543"
   /translation="MSFLRGIIDFSSILNEESKDDPSVSSSTSSSESNIGDVPVT
NERIAYLKGYPDLAKEIAGVRAEWMGLHDDALLHYNAQRINNEATSPSPSYIS
SEKEXVRSYRKISQNWQVSEIRLQALCVGSENKRTVAYPSSASVSTSRVKTLL
SOKTPVARGVATPRNPKDAAASPRKESGVNVDKLVEMINTTIVDRSPSVKWDVV
AGLNGAKQALLEWVILPAKRDLPFGLEPARVTSILGLLFGPGNGKTKLAKAVAS
EQATPENVASASLTKSDISIMSTRSENEASRLKSELIOPDGVTSNPDVLVII
IGATNPQELDDAVLRVLRIYVLPDSNVAKLFLTKLCKQPSLSDGIDIKLVE
TEGVLKCIKCKHFISQVTDKXSGSLQALCEEAAMWPIRELGANILITIQANKVLN
FSQVQINVRVCSLKLGLINRCVHVTKIST1"
6429..8288
   /gene="At2g45510"
   /note="synonym: F17K2.4"
   join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
   /gene="At2g45510"
   join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
   /gene="At2g45510"
   /codon_start=1
   /product="putative cytochrome P450"
   /protein_id="AAC06153.1"
   /db_xref="GI:2979544"
   /translation="NEILTSIAITVATTIVFVLCFTIYIMIRITFGKSRNDRKYPVH
ATVFDLLHSDSLYDYEIAAREKTYRFLSPQSEILTADPNVHEHLKTRFONYSK
GHSSRENMDLLGHYGFVADGKWKQKLSGFSFSTRVLRDPSCSVFRRNASKLVGF
VSEFALSGKAFDAQDLMLRCTLDSIPKVGFGVELKLDGFSKEGOFWEAFDEGNVAT
SRFIDPLWKLKWFNIGSQSKLSIATIDKFTYSLITTKRELAKQNTVWRDIL
SRFLVSEKDPENMDKYLRIILNFMIAKGDTTAAALSFWLMLCKNPLVQEKLVQE
IRDVTSHEKTTDVTYVINGDAEPEKPERLKDGLFPPVPMKCAENDDV
LPDGHVSVSGDNIYYIAYANGMTYWGQDAEPEKPERLKDGLFPPVPMKCAENDDV
GPRICLGKDPAYQMKIVSMALLHFFRFKMDENSKVYVKRMILTLHVDGGLHLCAIPR
TST"
   /complement(8412..9632)
   /gene="At2g45520"
   /note="synonym: F17K2.5; supported by full length cDNA:
Ceres:8220"
   /complement(join(<8412..8678,8768..8903,8987..9120,
9289..9356,9459..9632))
   /gene="At2g45520"
   /complement(join(8553..8678,8768..8903,8987..9120,
9289..9356,9459..9612))
   /gene="At2g45520"
   /codon_start=1
   /product="expressed protein"
   /protein_id="AAC06155.1"
   /db_xref="GI:2979546"
   /translation="WGGKGRERKNYLAHGGPARLPBPDRSKQDDVPSTLRIMN
YTSFPHDSTKQVVEKKEKLAEVVVVATPSDGDSDSVSEKKKKRKNQMTDLRFEN
ELAIDGRSKRKKKKYWEAKKKKNGKGTDLREPNFPKHEQIFGDDVVQAPDLKA
VVPKARKTISASQERLQALDAYRSRKGWTFPGVPIPAVMQ"
9408..9427
   /tpt_family="(CA)n"
   /tpt_family="(9578..9632)
   /tpt_family="(GA)n"
9822..10798
   /gene="At2g45530"
   /note="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"
9822..10798
   /gene="At2g45530"
9948..10670
   /gene="At2g45530"
   /codon_start=1
   /product="expressed protein"
   /protein_id="AAC06154.1"

```

74411 ATCGCTGGTCCGTCGAATCAGCCTCAGTTATTAGCTTTTCCCGCGCGCTGCTTCGGCG 74355

781 TCGTCTTTACGTCCGCGCTGTTTCAACAGAGCTTCCACGATGGCTAGACACACCTCTTACAA 840

74351 TCGTCTTTACGTCCGCGCTGTTTCAACAGAGCTTCCACGATGGCTAGACACACCTCTTACAA 74292

841 GTGTCTTCAACGAGCGGCTTTGTATCCGTTTCAGAGCTTAGCGGTTGGAATTTATCAAGA 900

74291 GTTGTCTTCAACGAGCGGCTTTGTATCCGTTTCAGAGCTTAGCGGTTGGAATTTATCAAGA 74232

901 GCGAGCTCGGTTTATGGCTCCGAGCTCAAGCTCAGCGGTAACAACCGGTAGTTCATCGTCA 960

74231 GCGAGCTCGGTTTATGGCTCCGAGCTCAAGCTCAGCGGTAACAACCGGTAGTTCATCGTCA 74172

961 ATTGCAACAAACGAGCGGCTTTGTATCCGTTTCAGAGCTTAGCGGTTGGAATTTATCAAGA 1020

74171 ATTGCAACAAACGAGCGGCTTTGTATCCGTTTCAGAGCTTAGCGGTTGGAATTTATCAAGA 74112

1021 GAGCTTCAACGAGCTTCAATGAGCACCACAGCAGCGGTCATCGAACCACTGA 1071

74111 GAGCTTCAACGAGCTTCAATGAGCACCACAGCAGCGGTCATCGAACCACTGA 74061

BT008779 882 bp mRNA linear PLN 13-JUN-2003

Arabidopsis thaliana At5g51910 mRNA, complete cds

BT008779.1 GI:31711723

FLI CDNA.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota, Viridiplantae

Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots,

rosids, eurosids II, Brassicales, Brassicaceae, Arabidopsis

1 (bases 1 to 882)

Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Bowser,L., Carninci,P., Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.

Arabidopsis ORF clones

Unpublished

2 (bases 1 to 882)

Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Bowser,L., Carninci,P., Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.

Direct Submission

Submitted (13-JUN-2003) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : RIKEN Arabidopsis Full-length cDNA) : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the pUMI (ORF) clones using the RAFL cDNAs: Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Bowser,L., Chan,M.M., Chang,C.M., Dale,J.M., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Davis,R.W., Theologis,A., and Ecker,J.R.

Kim,C.J. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to

this work. Shinozaki K. (RIKEN GSC) and Ecker, J.R. (ssp/salk) contributed equally to this work as PIs.

```

FEATURES
  source
    1..882
      /organism="Arabidopsis thaliana"
      /mol_type="mRNA"
      /db_xref="taxon:3702"
      /chromosome="5"
      /clone="U60996"
      /notes="This clone is in pUNI 51"
      ecotype: Columbia
    1..882
      /notes="unknown protein"
      /codon_start=1
      /product="At5g51910"
      /protein_id="AAP68218.1"
      /db_xref="GI:31711724"
      /translation="MESNHGNAIOVDVTTWTHLSDNPDKPKGMMLKQEDGVLY
      PVKTKPAKPKTSKDRHTKVEGRIRMPAGCAARVOLTRELGHKSGETIRWILLE
      RAEPALIEATGTGTVPAIAVSNGTLKIPITSSVLNDGGDGDGLIKRRKRNCTSD
      FVDVNSCHSVTSGLAIPITASNYGNILNVNTQGFVPFPMGNGTAFVTGGPDQMGQ
      MWAIPVTATPFLNVGARPVSVSYNSADAEEMETSGGGTTQPLRDFSLIYDKREL
      QFLGSGNSPSSCHET"
  CDS
    236 a 192 c 250 g 204 t
      BASE COUNT
      ORIGIN
        Query Match 14.7%; Score 157.8; DB 8; Length 882;
        Best Local Similarity 75.9%; Pred. No. 1.4e-36;
        Matches 195; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 195 GACAGGACCACTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAGGTTTCAAGGAAG 254
DB 141 GACTTAACCGGCTCCGAGAGACGACTTTTAAGACCGTCACACGAGGTTAGAGAGCG 200
QY 255 AGGAGAAGGATACGAGTGCCTCCACGCTGCGCGTACGATTTTCAATTAACCTCGAGA 314
DB 201 AGGTCCGAGGATCGCAATGCCGCGGGTTCGCTCGGGTCTTTCAATTAACCGTCA 260
QY 315 GTTAGGTACAAATCCGACCGGAGAAAGTTCGGTGGTTCGAGACCGTGCAGCGCGC 374
DB 261 ACTTGGGTACAAATCCGACCGGAGAAAGTTCGGTGGTTCGAGACCGTGCAGCGCGC 320
QY 375 GATTATAGCCGCCACGGGTACGGGAACGGTTCGCCCATCGCCATGTCGGTTTAAACGGAAC 434
DB 321 GATTAATTGAACCAACCGGAACCGGAACTGTACCGGCTATTGCTGTATCGGTTAAACGGAAC 380
QY 435 CTTAAAAATCCCGACGA 451
DB 381 TTTAAAAATCCCGACGA 397

RESULT 7
AK118579 1046 bp mRNA linear PLN 06-DEC-2002
LOCUS Arabidopsis thaliana At5g51910 mRNA for unknown protein, complete cds, clone: RAFL19-81-E01.
DEFINITION Arabidopsis thaliana full-length cDNA
ACCESSION AK118579
VERSION AK118579.1 GI:26452181
KEYWORDS FLI CDNA; CAP trapper.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
  AUTHORS Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.
  TITLE Arabidopsis thaliana full-length cDNA
  JOURNAL Published Only in Database (2002)
  REFERENCE 2 (bases 1 to 1046)
  AUTHORS Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P.,

```

```

Kawai, J., Hayashizaki, Y. and Shinozaki, K.
Direct Submission
Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences
Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa
230-0045, Japan (E-mail: msei@gs.riken.go.jp,
URL: http://bigweb.gsc.riken.go.jp, Tel: 81-45-503-9625,
Fax: 81-45-503-9586)
COMMENT
  An Arabidopsis full-length cDNA library was constructed essentially
  as reported previously (Seki et al. (1998) Plant J. 15:707-720;
  Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI
  and XbaI was ligated to modified Lambda phage-1-E vector (Carninci et
  al. (2001) Genomics 77:79-90) digested with BamHI and SalI.
  This clone is in a modified pBluescript vector.
  Please visit our web site (http://bigweb.gsc.riken.go.jp/) for
  further details.
  Location/Qualifiers
    1..1046
      /organism="Arabidopsis thaliana"
      /mol_type="mRNA"
      /cultiivar="Columbia"
      /db_xref="taxon:3702"
      /chromosome="5"
      /clone="RAFL19-81-E01"
      /note="common name: thale cress"
    1..1046
      /gene="At5g51910/MJM18.6"
      /CDS
        47..928
          /gene="At5g51910/MJM18.6"
          /codon_start=1
          /product="unknown protein"
          /protein_id="BAC43179.1"
          /db_xref="GI:26452182"
          /translation="MESNHGNAIOVDVTTWTHLSDNPDKPKGMMLKQEDGVLY
          PVKTKPAKPKTSKDRHTKVEGRIRMPAGCAARVOLTRELGHKSGETIRWILLE
          RAEPALIEATGTGTVPAIAVSNGTLKIPITSSVLNDGGDGDGLIKRRKRNCTSD
          FVDVNSCHSVTSGLAIPITASNYGNILNVNTQGFVPFPMGNGTAFVTGGPDQMGQ
          MWAIPVTATPFLNVGARPVSVSYNSADAEEMETSGGGTTQPLRDFSLIYDKREL
          QFLGSGNSPSSCHET"
  BASE COUNT 277 a 208 c 287 g 274 t
  ORIGIN
    Query Match 14.7%; Score 157.8; DB 8; Length 1046;
    Best Local Similarity 75.9%; Pred. No. 1.4e-36;
    Matches 195; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 195 GACAGGACCACTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAGGTTTCAAGGAAG 254
DB 187 GACTTAACCGGCTCCGAGAGACGACTTTTAAGACCGTCACACGAGGTTAGAGAGCG 246
QY 255 AGGAGAAGGATACGAGTGCCTCCACGCTGCGCGTACGATTTTCAATTAACCTCGAGA 314
DB 247 AGGTCCGAGGATCGCAATGCCGCGGGTTCGCTCGGGTCTTTCAATTAACCGTGA 306
QY 315 GTTAGGTACAAATCCGACCGGAGAAAGTTCGGTGGTTCGAGACCGTGCAGCGCGC 374
DB 307 ACTTGGGTACAAATCCGACCGGAGAAAGTTCGGTGGTTCGAGACCGTGCAGCGCGC 366
QY 375 GATTATAGCCGCCACCGGTTACGGGAACGGTTCGCCCATCGCCATGTCGGTTTAAACGGAAC 434
DB 367 GATTAATTGAACCAACCGGAACCGGAACGTGTACCGGCTATTGCTGTATCGGTTAAACGGAAC 426
QY 435 CTTAAAAATCCCGACGA 451
DB 427 TTTAAAAATCCCGACGA 443

RESULT 8
AY086379 1156 bp mRNA linear PLN 14-APR-2003
LOCUS Arabidopsis thaliana clone 247645 mRNA, complete sequence.
DEFINITION Arabidopsis thaliana full-length cDNA
ACCESSION AY086379
VERSION AY086379.1 GI:21405089
KEYWORDS FLI CDNA.

```


[illegible]

354 CCGACGCTAAGCCGACCGCTTCAACGACCGTACGAAAGACCGACACACGGAAGCTAGA 413

249 AGAAGAGGAGAGATACGGATCGCTCCACGGTGTGGGTAGATTTTCAATTAAAC 308
 414 CGAAGAGGAGAGAGATTAAGATGCGCGGATATGTGCAGCTAGGGTTTTCAGCTAAC 473
 309 TCGAGAGTTAGGTCAAAATCCGACGGCGAAGACGATTCGGTGGTGTGTGGAGAACGCTGA 368
 474 GCGAGAGCTAGGTCAATAATCCGACGGTGAGACAATAGAGTGGCTTCTTCAACAGCTGA 533
 369 GCGCGGATATAGCGGCCACCGGTACGGGAACGGTTCGCCCATCGCCATGCGGTTAA 428
 534 ACCATCTGAATCGCGGCCACCGGAACCGGAACAATCCGGCGAATTCACCTCTTTAAA 593
 429 C 429
 594 C 594

RESULT 12
 ATEP2 101154 bp DNA linear PLN 30-NOV-1999
 Arabidopsis thaliana DNA chromosome 3, BAC clone FIP2.
 AL132955
 AL132955.1 GI:6434210
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1
 Choisne.N., Robert,C., Brottier,P., Wincker,P., Cattolico,L.,
 Artiguenave,F., Saurin,W., Weissenbach,J., Mewes,H.W., Lemcke,K.,
 Mayer,K.F.X., Quetier,F. and Salanoubat,M.
 Unpublished
 2 (bases 1 to 101154)
 EU Arabidopsis sequencing,project.
 Direct Submission
 Submitted (30-NOV-1999) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitzz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,project
 Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
 d'interet Public, Centre National de Sequenace - GENOSCOPE; 2 rue
 Gaston Cremieux, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.

FEATURES
 source
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 113..6526
 /gene="FIP2.10"
 join(113..430,519..1199,1277..1420,1498..1788,1869..2054,
 2135..2290,2378..2437,2672..2734,3123..3214,3344..3440,
 3916..4032,4174..4392,4519..4659,4737..4904,5138..5322,
 5491..5584,5671..5759,5842..6067,6228..6302,6413..6526)
 /gene="FIP2.10"
 XCAP-E - strong similarity to CHROMOSOME ASSEMBLY PROTEIN
 XCAP-E - Xenopus laevis, SWISSPROT:XCAP_XENLA"
 /codon_start=1
 /product="chromosome assembly protein homolog"
 /protein_id="CAB61972.1"
 /db_xref="GI:6522529"
 /translation="WHIKICLEGKSYATRTTVVGFPHFNALITGLSGSKNILD
 ICFVLGITNQVLRANLQELVYKQAGATITVATSVTFDSENRNRSLIGHEDHSKIT
 VTQIVVGGNKYLKLAQNVQNFHSVLQNNPHFLIMQGRITKVLNKKPME
 ILSMGEAAGTMYENKEAAALTKKQTKYDELINKLEKLDILPALEKIRKESQYM
 QWANGAEIDRLKRFVCAVEYVQAKIRDSIHVVVEMKIKMTGIDETDKTQGEISE

intron 3441. .3915
/gene="FIP2.10"
/number=10
exon 3916. .4032
/gene="FIP2.10"
/number=11
intron 4033. .4173
/gene="FIP2.10"
/number=11
exon 4174. .4392
/gene="FIP2.10"
/number=12
intron 4393. .4518
/gene="FIP2.10"
/number=12
exon 4519. .4659
/gene="FIP2.10"
/number=13
intron 4660. .4736
/gene="FIP2.10"
/number=13
exon 4737. .4904
/gene="FIP2.10"
/number=14
intron 4905. .5137
/gene="FIP2.10"
/number=14
exon 5138. .5322
/gene="FIP2.10"
/number=15
intron 5323. .5490
/gene="FIP2.10"
/number=15
exon 5491. .5584
/gene="FIP2.10"
/number=16
intron 5585. .5670
/gene="FIP2.10"
/number=16
exon 5671. .5759
/gene="FIP2.10"
/number=17
intron 5760. .5841
/gene="FIP2.10"
/number=17
exon 5842. .6067
/gene="FIP2.10"
/number=18
intron 6068. .6227
/gene="FIP2.10"
/number=18
exon 6228. .6302
/gene="FIP2.10"
/number=19
intron 6303. .6412
/gene="FIP2.10"
/number=19
exon 6413. .6526
/gene="FIP2.10"
/number=20
gene 6370. .8121
/gene="FIP2.20"
gene complement(join(6970. .7100,7191. .7457,7537. .7685,
7913. .8121))
/gene="FIP2.20"
CDS complement(join(6970. .7100,7191. .7457,7537. .7685,
7913. .8121))
/gene="FIP2.20"
/note="strong similarity to CHLOROPHYLL A-B BINDING
PROTEIN 4 PRECURSOR (LHCI TYPEIII CAB-4) - Arabidopsis
thaliana, SWISSPROT:CB24_ARATH"
/codon_start=1
/product="CHLOROPHYLL A-B BINDING PROTEIN 4 PRECURSOR

homolog"

Query Match 12.0%; Score 129; DB 8; Length 101154;
Best Local Similarity 71.0%; Pred. No. 1.9e-27;
Matches 171; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 189 GTCTTCGACGAGGACCACTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAAGTTGA 248
DB 72666 GGCAGCTAAAGACCCCGTTGAACGAGGTCGACGAAGACCGACACCGAAGATGA 72725
QY 249 AGGAAGAGGAGAGGATACGGATGCTCCACGTTGCGGCTAGGATTTTCAATTAC 308
DB 72726 CGGAAGAGGAGGAGATAAGGATCGCGGCTTATGTGCAGCTAGGTTTTTCAAGTAA 72785
QY 309 TCGAGAGTTCAGTCAAAATCCGACGCGGCAAGGATTCGTTGGTTCGAGAAAGCTGA 368
DB 72786 GCGAGAGCTAGGTCAATAATCCGACGCTGAGACAATAGATGGCTTCITTCACAAAGCTGA 72845
QY 369 GCGGCGGATTATAGCGCCACGCGGTACGCGAAGCGTTCCCGCCATCGCCATTCGTTAA 428
DB 72846 ACCATCTGTAATCGCGCCACCGAACCGAACAATCCCGGCAATTCACCTTCITTA 72905
QY 429 C 429
DB 72906 C 72906

RESULT 13
BX322174
LOCUS
DEFINITION Arabidopsis thaliana transposon insertion STS SM_3.35210, sequence
tagged site.
ACCESSION BX322174
VERSION BX322174.1 GI:29420305
KEYWORDS STS; STS, sequence tagged site.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1
AUTHORS Clarke, J.H., Bowles, B., Carter, J., Hart, D., McCullagh, B.,
Murphy, G., Langham, S., LeGrys, C., Jones, J.D.G. and Bevan, M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 451)
AUTHORS Clarke, J.H.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2003) Clarke J.H., John Innes Centre, Colney
Lane, Norwich, NR4 7UU, UK
COMMENT AT denotes an activation tag dissociation transposon within a
single line, ET an enhancer trap dissociation transposon, GT a gene
trap dissociation transposon, MT a mis-expression enhancer trap
dissociation transposon, SM a defective suppressor mutator
transposon, _3 denotes a sequence derived from the 3' end of the
transposon, _5 denotes a sequence derived from the 5' end of the
transposon BBSRC GARNET, ARIS project
On-line seed stock requests: <http://nasc.nott.ac.uk/> NASC stock
code: N121921.

FEATURES
source
1..451
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia-0 NASC stock code N1092"
/db_xref="taxon:3702"
/clone="Al132955"
/notes="Derived from superpool 26.24 NASC code Unknown"
1..451
/standard_name="SM_3.35210"
142 a 139 c 87 g 83 t

STS
BASE COUNT 142 a 139 c 87 g 83 t
ORIGIN

Query Match 11.6%; Score 124.6; DB 11; Length 451;
Best Local Similarity 73.1%; Pred. No. 2e-26;

Matches 160; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

2y 189 GTCTTCAGAGGACCACTTGAAGAGAGCTTCGACTAAGACCTGCACAGAGTTGA 248

3b 233 GCAGCTAAAGACCCAGCTTGAACAGAGCTTCGAGGAGGACCGACACAGAAAGTAGA 292

2y 249 AGGAAGAGGAGAGGATACGATGCTCCAGCTGCTCGGCTAGGATTTTCAATTAAAC 308

3b 293 CGGAAGAGGAGGAGGATTAAGGATGCGCGGTTATGTGACGTAGGTTTTCACCTAAC 352

2y 309 TCAGAGTAGTTCACAAATCCAGCGGAGAAAGATTCGTTGTTGCGAGAACGCTGA 368

3b 353 GCAGAGAGTAGGTCAATAATCCAGCGGTGAGCAATAGAGTGGCTTCTCAACAAGCTGA 412

2y 369 GCGCGGATTTATAGCGCGCACGGGTACGGGAACGGTTCC 407

3b 413 ACATCTGTAATCGCGCCACCGGACCGGACCAATCCC 451

RESULT 14
 LOCUS AY081706 978 bp mRNA linear PLN 18-MAR-2002
 DEFINITION Arabidopsis thaliana At1g69690/T6C23_11 mRNA, complete cds.

ACCESSION AY081706.1 GI:19547990

VERSION Arabidopsis thaliana (thale cress)

KEYWORDS Arabidopsis thaliana

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana (thale cress)

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

SPERMATOPHYTES; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 978)

Chen, R., Chen, H., Kim, C.J., Meyers, M.C., Ban, J., Bowser, L.,

Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,

Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,

Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,

Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,

Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K.,

Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A.,

and Ecker, J.R.

Arabidopsis ORF clones

Unpublished

2 (bases 1 to 978)

Chen, R., Chen, H., Kim, C.J., Meyers, M.C., Ban, J., Bowser, L.,

Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,

Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,

Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,

Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,

Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K.,

Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A.,

and Ecker, J.R.

Direct Submission

Submitted (28-FEB-2002) Salk Institute Genomic Analysis Laboratory

(SIGNAL), Plant Biology Laboratory, The Salk Institute for

Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,

USA

RIKEN Genomic Sciences Center (GSC) members carried out the

collection and clustering of RAPL cDNAs (RAPL cDNA : RIKEN

Arabidopsis Full-length cDNA) : Seki, M., Narusaka, M., Ishida, J.,

Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,

Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the

sequencing and annotation of the RAPL cDNAs: Chen, R., Chen, H.,

Kim, C.J., Meyers, M.C., Shinn, P., Ban, J., Bowser, L., Chan, M.M.,

Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Jones, T.,

Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M.,

Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C.,

Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Davis, R.W.,

Theologis, A., and Ecker, J.R.

Chen, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to

this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)

contributed equally to this work as PIs.

Location/Qualifiers

1..978

/organism="Arabidopsis thaliana"

/molecule="mRNA"

/db_xref="taxon:3702"

/chromosome="1"

/clone="U14268"

/notes="This clone is in PUNI 51"

ecotype: Columbia"

1..978

/notes="unknown protein"

/codon_start=1

/product="At1g69690/T6C23_11"

/protein_id="AAL87359.1"

/db_xref="GI:19547991"

/translation="MDPDPHNRPNFPLQLDSSSTSSSTSLAISTSEPNSEPKK

PPKRTSTKDRHTKVEGRGRIRMPACARVOLTRELGHKSDGETTEMLIQAEPA

VIAATGTGTPANFTSLNISLSSSSLSAAHLRTTPSYVPHSPQSMTHLQHQHPA

VRPNESHSSSSSSQLLDHNMGNLYVOSTAGSLPTSQSPATAPFWSSGNDNTNLWA

FNINPHGVAGVADYVNPNSGGSGGVHLMNFAPIALPFGQPLASGYGGGGGGG

EHSHYGLAALNAARVPAETGNHNNQNRDGHNNHNDGDSSTSHS"

BASE COUNT 245 a 269 c 201 g 263 t

ORIGIN

Query Match 11.6%; Score 124.6; DB 8; Length 978;

Best Local Similarity 65.0%; Pred. NO. 2.2e-26;

Matches 184; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 202 CCACCATTTGAAGAGAGCTTCGACTAAGACCGCTCACACGAGGTTGAAGAGAGGAGA 261

DB 133 CCTCTCTTAACGAACTCTACTAAGACCGACACACCAAGTCGAGCGCGCCGT 192

QY 262 AGGATAGGATGCTGCCAGTGTGGGTAGATTTTCAATTAATCGAGATGAGT 321

DB 193 CGGATCGGTATGCTGCTCCATGTGTGTCAGCTGCTTTTCACTACACGTGAGT 252

QY 322 CACAAATCCACGCGGAGAAAGATTCGTTGTTGAGAGAAAGCTGAGCGCGGATATA 381

DB 253 CACAAATCCGAGGTTGAATTTGATGCTACTCAACAGCAGACACGCGGTTATA 312

QY 382 GCGCCACCGGTACGGAGACGGTTCCGCCATCGCATTCGTTTAAACGACCTTAAA 441

DB 313 GCGCTACAGGAGTGGACCATTCGCGCTTAACCTTCTTTTAAACATCTCACTCGT 372

QY 442 ATCCCGACGAGCAGGAGCGTGTCTTGATATGGTGAAATC 484

DB 373 AGCTCAAGATCTTCTCTCTGCTGCTCACTTCCTGACACTC 415

RESULT 15

AY056384 1440 bp mRNA linear PLN 03-OCT-2001

LOCUS Arabidopsis thaliana At1g69690/T6C23_11 mRNA, complete cds.

DEFINITION Arabidopsis thaliana

ACCESSION AY056384

VERSION AY056384.1 GI:15912212

KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 1440).

Chen, R., Chen, H., Kim, C.J., Koesena, E., Meyers, M.C., Ban, J.,

Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,

Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G.,

Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,

Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K.,

Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,

Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K.,

Davis, R.W., Theologis, A. and Ecker, J.R.

Arabidopsis cDNA clones

Unpublished

TITLE

JOURNAL

REFERENCE
AUTHORS

2 (bases 1 to 1440)
 Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J.,
 Bowers, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,
 Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,
 Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K.,
 Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,
 Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K.,
 Davis, R.W., Theologis, A. and Ecker, J.R.
 Direct Submission
 Submitted (10-SEP-2001) Salk Institute Genomic Analysis Laboratory
 (SIGAL), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
 USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
 Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Cheuk, R., Chen, H.,
 Kim, C.J., Koesema, E., Meyers, M.C., Shinn, P., Banh, J., Bowers, L.,
 Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G.,
 Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,
 Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A.,
 Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,
 Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
 contributed equally to this work as PIs.

FEATURES

source

Location/Qualifiers
 1..1440
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="RAFL09-32-L20(R14268)"
 /note="ecotype: Columbia"
 1..9

5'UTR
CDS

10..987
 /note="unknown protein"
 /codon_start=1
 /product="At1g89690/T6C23_11"
 /protein_id="AAL08240.1"
 /db_xref="GI:15912213"
 /translation="MDPPDHNRPFLQLDSSSTSLAIITSEPNSEPK
 PPKETSTKDRHTKVEGRIRMPACARVOLTRELGHKSDGTIEWLLQOAEPA
 VIAATGTIPANFTSLAIRSRSLRAHLRTTSSVYFHSQSMTHLOHOHO
 VPKNESHSSSQLLDHOMGYLVQSTLPTSPATAPFWSSGDNTQNLWA
 FNINPHSGVAVGVYNFNSGGSGGVHLMNFAFIALFSCGPLASGIGGGGGGG
 EHSHYGVLAALNAAFRPVAETGNHNNQNRDGDHNNHNDGDSSTSHS"
 988..1440

3'UTR
BASE COUNT
ORIGIN

383 a 320 c 264 g 473 t

Query Match 11.6%; Score 124.6; DB 8; Length 1440;
 Best Local Similarity 65.0%; Pred. No. 2.3e-26;
 Matches 184; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
 y 202 CCACCATTCAGAGAGCTTCGACTAAGAACCGTCAACGAGGTGAGGAGGAGGAGA 261
 b 142 CTTCTCTCTAAGACCACTTCTACTAAGACCGGACACACCAAGTCGAAGCCGAGCCGT 201
 y 262 AGGATACGATGCTGCGACGTCGCGCTAGGATTTTCAATTAACTCGAGAGTTAGGT 321
 b 202 CGATCCGATATGCTTGCCTATGTGTGCTGACGCTGCTTTTCACTCACAGTGGTGGT 261
 y 322 CACAAATCCGACGGCGAAACGATTCGTTGTTGGAGAACGCTGAGCCGCGGATATA 381
 b 262 CACAAATCCGACGGGTGAATCTATTGAGTGGCTACTCCAAAGCAGAACACGCGGTTATA 321

Qy 382 GCCGCCACGGGTACGGAGACGGTTCCGCCATCGCCATCGGTTACGGTAACGGAACCTTAAA 441
 Db 322 GCCCTACAGGACTGGAACCAATCCGGGTAACTTCACTTCTTTAAACATCTCACTTCGT 381
 Qy 442 ATCCCGACGACGACGAAACGCTGATTCTGATATGGGTGAAATC 484
 Db 382 AGCTCAAGATCTTCTCTCTCTGCTGCTCATCTTCGTGACAACTC 424

Search completed: February 2, 2004, 00:52:51
 J05 time : 4071.41 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:36:05 ; Search time 313.739 Seconds
(without alignment)
9214.976 Million cell updates/sec

Title: US-09-938-842A-1034

Perfect score: 1071

Sequence: 1 atggcgacaattcgaagct.....cacggctcatgaaccactga 1071

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_13Jun03.*
1: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1071	100.0	1071	24	Arabidopsis thalia
2	448	41.8	460	25	Arabidopsis thalia
3	439	41.0	453	24	Arabidopsis thalia
4	157.8	14.7	1156	21	Arabidopsis thalia
5	129	12.0	1728	24	Arabidopsis thalia
6	124.6	11.6	1231	24	Arabidopsis thalia
7	124.6	11.6	1297	21	Arabidopsis thalia
8	118	11.0	1081	21	A. thaliana VDBP

9	111.2	10.4	422	25	ABX62217	Arabidopsis thalia
10	111.2	10.4	1009	24	ABK65215	Arabidopsis thalia
11	111.2	10.4	1486	21	AAC33900	Arabidopsis thalia
12	100	9.3	1172	24	ABK65353	Arabidopsis thalia
13	98.6	9.2	1263	24	ABK65216	Arabidopsis thalia
14	93.2	8.7	1379	21	AAC44188	Arabidopsis thalia
15	89	8.3	83698	24	ABN85167	Arabidopsis yellow
16	88.8	8.3	750	21	AAC53759	Arabidopsis thalia
17	79.6	7.4	419	25	ABX22048	Human GDP-mannose
18	43.6	4.1	447	24	ABL93996	Arabidopsis thalia
19	36	3.4	2420	23	ABL17625	Drosophila melanog
20	35.8	3.3	6161	23	ABL09562	Drosophila melanog
21	34.6	3.2	1062	12	AQ11001	BanI restriction e
22	34.2	3.2	8208	23	ABL17624	Drosophila melanog
23	34	3.2	554	18	AAV74924	Staphylococcus aur
24	33.6	3.1	463	13	AAQ26483	Recombinant bovine
25	33.6	3.1	463	17	AAT35894	Bovine (Ala47, Gly9
26	33.6	3.1	3436	23	ABL03290	Drosophila melanog
27	33.4	3.1	888	22	AAI94426	Human neuroblastom
28	33.4	3.1	1644	23	ABL16335	Drosophila melanog
29	33.4	3.1	4882	23	ABL01868	Drosophila melanog
30	33.4	3.1	5675	23	ABL16334	Drosophila melanog
31	33.2	3.1	386	22	AAF65986	Novel human polyu
32	33.2	3.1	469	22	ABA45750	Human breast cell
33	33.2	3.1	469	22	ABA56260	Human foetal liver
34	33.2	3.1	469	22	ABA25899	Probe #4365 for ge
35	33.2	3.1	469	22	AAK04443	Human brain expres
36	33.2	3.1	469	22	AAK29936	Human bone marrow
37	33.2	3.1	469	22	AAI14529	Probe #4462 for ge
38	33.2	3.1	469	22	AAI35902	Probe #4588 used t
39	33.2	3.1	469	22	AAI04351	Probe #4342 used t
40	33.2	3.1	469	23	ABS29588	Human liver single
41	33.2	3.1	469	24	ABS04509	Human genome-dariv
42	33.2	3.1	1893	25	ABT21247	Aspergillus fumiga
43	33.2	3.1	10732	21	AAA10594	Gene encoding a su
44	33	3.1	3011208	24	ABQ69245	Listeria innocua D
45	32.8	3.1	4403765	22	AAI99683	Mycobacterium tube

ALIGNMENTS

RESULT 1	ABZ13229	standard, DNA; 1071 BP.
ID	ABZ13229	standard, DNA; 1071 BP.
XX	XX	
AC	ABZ13229	
XX	XX	
DT	21-JAN-2003	(First entry)
XX	XX	
DE	Arabidopsis thaliana stress regulated gene SEQ ID NO 1034.	
XX	XX	
KW	Arabidopsis thaliana; plant; gene; stress; transgenic; ds.	
XX	XX	
OS	Arabidopsis thaliana.	
XX	XX	
PN	WO20021655-A2	
XX	XX	
PD	28-FEB-2002	
XX	XX	
PF	24-AUG-2001	2001WO-US26585.
XX	XX	
PR	24-AUG-2000	2000US-227866P.
PR	26-JAN-2001	2001US-264647P.
PR	22-JUN-2001	2001US-300111P.
XX	XX	
PA	(SCRI) SCIPPS RES INST	
PA	(SYGN) SYNGENTA PARTICIPATIONS AG.	
XX	XX	
PI	Harper JF, Krops J, Wang X, Zhu T,	
XX	XX	
DR	WPI; 2002-304127/34.	
XX	XX	

QY 607 CTGGCATCATCACCCTGGCTCGCAACTCTCTGCGCAAGGAATGATCCGATGTGGCT 666
Db 453 CTGGCATCATCACCCTGGCTCGCAACTCTCTGCGCAAGGAATGATCCGATGTGGCT 394
QY 667 ATTCCATCAACCGCAATGATCCGACGGTCGCGAGCTTTCTTCTTGATTCACAAATCGCT 726
Db 393 ATTCCATCAACCGCAATGATCCGACGGTCGCGAGCTTTCTTCTTGATTCACAAATCGCT 334
QY 727 GGTCCGTGCAATCAGCCTCAGTTATAGCTTTTCCGCGCGCGGTCTTCCGCGTCTTCCGCGTCT 786
Db 333 GGTCCGTGCAATCAGCCTCAGTTATAGCTTTTCCGCGCGCGGTCTTCCGCGTCTTCCGCGTCT 275
QY 787 TACGTCGCGCTGTTCAACAGGCTTCCAGATGGGTAGACCACTCTTTCACAGTTGTT 846
Db 274 TACGTCGCGCTGTTCAACAGGCTTCCAGATGGGTAGACCACTCTTTCACAGTTGTT 215
QY 847 CCAACAGCGGCTTTGTTATCCGTTTCCAGCTTAGCGGTTCGAATTTATCAAGAGCGACG 906
Db 214 CCAACAGCGGCTTTGTTATCCGTTTCCAGCTTAGCGGTTCGAATTTATCAAGAGCGACG 155
QY 907 TCGGTATGGCTCCAGCTCAGCTCAGCGGTAAACCGGTAGTTCATCGTCAATTGCA 966
Db 154 TCGGTATGGCTCCAGCTCAGCTCAGCGGTAAACCGGTAGTTCATCGTCAATTGCA 95
QY 967 ACAACAGCGCACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAAACAGAGCTT 1026
Db 94 ACAACAGCGCACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAAACAGAGCTT 35
QY 1027 CACGAGTTATGAGCACCAACAGCAGCGTTCAT 1060
Db 34 CACGAGTTATGAGCACCAACAGCAGCGTTCAT 1

RESULT 4

AAC39595

ID AAC39595 standard; DNA; 1156 BP.

XX AC AAC39595;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 25210.

XX KW Hybridisation assay; genetic mapping; gene expression control;

XX KW protein identification; signal transduction pathway;

XX KW metabolic pathway; promoter; termination sequence; ss.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 08-APR-1999; 99US-0128234.

XX PR 16-APR-1999; 99US-0128714.

XX PR 19-APR-1999; 99US-0129845.

XX PR 21-APR-1999; 99US-0130077.

XX PR 23-APR-1999; 99US-0130449.

XX PR 28-APR-1999; 99US-0130510.

XX PR 30-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132407.

XX PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145085.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145276.
PR 23-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156658.
PR 29-SEP-1999; 99US-0156956.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.

PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.7%; Score 157.8; DB 21; Length 1156;
Best Local Similarity 75.9%; Pred. No. 4.9e-40;
Matches 195; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 195 GACAGGACCAATTAAGAGAGCTTCGACTAAAGACCGTCAACGAGGTTGAAGGAG 254
|||
Db 258 GACTAAACCGGCTCCGAAGAGAGACCGACTTCTAAAGACCGTCAACGAGGTTGAAGGAG 317
|||
QY 255 AGGAGAGAGTACGGATGCTGCCACGTGTCGGGTAGGATTTTCAATTAACTCGA 314
|||
Db 318 AGTCCGAGGATCCGAATCCGCGGGTTCGGTTCGGGTCTTTCAATTGACCGTGA 377
|||
QY 315 GTTAGGTCACAAAATCCGACGCGGAAACGATTCCGGTGGTTCGGGAAACGCTGAGCCGC 374
|||
Db 378 ACTTGGTCACAAATCCGACGCGGAAACGATACGTTATTGGAACGAGCTGAACCGGC 437
|||
QY 375 GATTATAGCCGCCACGGGTAGCGGAGCGGTTCCGCCCATCGCCATCGGTTAACCGAAC 434
|||
Db 438 GATAATTGAAGCAACCGGAACCGGAACTGTACCGGCTATTGCTGTATCGGTTAACCGAAC 497
|||
QY 435 CTTAAAAATCCCGACGA 451
|||
Db 498 TTTAAAAATCCCGACGA 514
|||

RESULT 5

ABK65233
ID ABK65233 standard; cDNA; 1728 BP.

AC ABK65233;

DT 02-JUL-2002 (first entry)

XX Arabidopsis cDNA encoding a transcription factor #85.

XX Plant; ss; gene; transcription factor; transgenic;
XX agriculture; metabolic chemical; environmental stress; drought;
XX microbial disease resistance; herbicide resistance; seed yield;
XX fruit yield; growth rate; leaf senescence; flower senescence.

OS Arabidopsis thaliana.

XX WO200215675-A1.

XX 28-FEB-2002.

XX 22-AUG-2001; 2001WO-US26189.

XX 22-AUG-2000; 2000US-227439P.

PR 16-NOV-2000; 2000US-0713994.

PR 16-APR-2001; 2001US-0837944.

XX (MEND-) MENDEL BIOTECHNOLOGY INC.

PA (PILG/) PILGRIM M.

PA (CREE/) CREELMAN R.

PA (DUBE/) DUBELL A J.

PA (HEAR//) HEARD J.
PA (JIAN//) JIANG C.
PA (KEDD//) KEDDIE J.
PA (ADAM//) ADAM L.
PA (RATC//) RATCLIFF O.
PA (REUB//) REUBER J L.
PA (RIEC//) RIECHMANN J L.
PA (YUGG//) YU G.
PA (PINE//) PINEDA O.

XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J;
PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX WPI; 2002-292022/33.
DR P-PSDB; RAU93047.

XX An isolated or recombinant polynucleotide used to produce a transgenic
PT plant -

XX Claim 4; Page 373-376; 941pp; English.

XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
XX encoding an Arabidopsis thaliana transcription factor, their variants,
XX complements, fragments, or related polynucleotide with 31% to 95%
XX sequence identity, where the plant possesses an altered trait as compared
XX to a wild-type or reference plant, or the plant exhibits an altered
XX phenotype as compared to a wild-type or reference plant, or the plant
XX exhibits ectopic expression or altered expression of one or more genes
XX associated with a plant trait as compared to a wild plant. Also included
XX are a transgenic plant comprising the polynucleotides, a computer
XX readable medium having stored sequence information, and identifying a
XX homologue sequence from a database comprising a plurality of known plant
XX sequences comprising inputting sequence information selected from one of
XX 464 fully defined sequences given in the specification. The isolated or
XX recombinant polynucleotide is used for producing a plant having a
XX modified trait, the method comprising selecting a polynucleotide that
XX encodes a polypeptide or an antisense nucleic acid, inserting the
XX polynucleotide or antisense nucleic acid into an expression vector,
XX introducing the vector into a plant or a cell of a plant to overexpress
XX the polypeptide or antisense nucleic acid, thereby producing a modified
XX plant, and selecting for a modified trait (e.g. increased
XX production of agriculturally useful proteins or metabolic chemicals,
XX pest tolerance, environmental stress response (e.g. drought), microbial
XX disease resistance, herbicide resistance, seed and fruit yield, growth
XX rate, leaf and flower senescence and many other traits listed in the
XX specification). The present sequence is one of the 232 polynucleotides
XX encoding an A. thaliana transcription factor.

XX Sequence 1728 BP; 526 A; 414 C; 362 G; 426 T; 0 other;

XX Query Match
XX Best Local Similarity 12.0%; Score 129; DB 24; Length 1728;
XX Matches 171; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 189 GTCTTCGACGAGCACCATTTGAGAGAGCTTCGACTTAAGACCGTCAACGAGGTGA 248
DB 420 GCGAGCTAAAGGACCCGTTGAAACGAGCGTCCGAGAAACGACACGACGAAAGTGA 479
QY 249 AGGAGAGGAGAGGATACCGATGCTGCTCCACCTGTGCGGTAGGATTTTCAATTAC 308
DB 480 CGGAAGAGGGAGGAGAAATGAGATGCGCGGTATGTGACACTAGGTTTTTACGTAC 539
QY 309 TCGAGAGTAGTGCACAAATCCGACGGCGAAACGATTGCGTGTGTGGAGACGCTGA 368
DB 540 GCGAGAGTAGTGCATAATCCGACGGTGAGACAAATAGATGGCTTTCTCAACAGCTGA 599
QY 369 GCGCGGATATAGCCGACGAGGTACGGACGGTTCCCGCATGCCCATGTCGTTAA 428
DB 600 ACCATCTGTAATCGCGCCACCGGAAACCGAACAATCCCGGAATTCCTCTTTAA 659
QY 429 C 429
DB 660 C 660

RESULT 6

ABK65217
ID ABK65217 standard; cDNA; 1231 BP.

XX AC ABK65217;

XX DT 02-JUL-2002 (first entry)

XX DE Arabidopsis cDNA encoding a transcription factor #69.

XX Plant; ss; gene; transcription factor; transgenic;
KW agriculture; metabolic chemical; environmental stress; drought;
KW microbial disease resistance; herbicide resistance; seed yield;
KW fruit yield; growth rate; leaf senescence; flower senescence.

XX OS Arabidopsis thaliana.

XX PN WO200215675-A1.

XX PD 28-FEB-2002.

XX PF 22-AUG-2001; 2001WO-US26189.

XX PR 22-AUG-2000; 2000US-227439P.

XX PR 16-NOV-2000; 2000US-0713394.

XX PR 16-APR-2001; 2001US-0837944.

XX (MEND-) MENDEL BIOTECHNOLOGY INC.

PA (PILG//) PILGRIM M.

PA (CREE//) CREELMAN R.

PA (DUBE//) DUBELL A J.

PA (HEAR//) HEARD J.

PA (JIAN//) JIANG C.

PA (KEDD//) KEDDIE J.

PA (ADAM//) ADAM L.

PA (RATC//) RATCLIFF O.

PA (REUB//) REUBER J L.

PA (RIEC//) RIECHMANN J L.

PA (YUGG//) YU G.

PA (PINE//) PINEDA O.

XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J;

PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;

XX WPI; 2002-292022/33.

DR P-PSDB; AAU93031.

XX An isolated or recombinant polynucleotide used to produce a transgenic
PT plant -

XX Claim 4; Page 315-317; 941pp; English.

XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
XX encoding an Arabidopsis thaliana transcription factor, their variants,
XX complements, fragments, or related polynucleotide with 31% to 95%
XX sequence identity, where the plant possesses an altered trait as compared
XX to a wild-type or reference plant, or the plant exhibits an altered
XX phenotype as compared to a wild-type or reference plant, or the plant
XX exhibits ectopic expression or altered expression of one or more genes
XX associated with a plant trait as compared to a wild plant. Also included
XX are a transgenic plant comprising the polynucleotides, a computer
XX readable medium having stored sequence information, and identifying a
XX homologue sequence from a database comprising a plurality of known plant
XX sequences comprising inputting sequence information selected from one of
XX 464 fully defined sequences given in the specification. The isolated or
XX recombinant polynucleotide is used for producing a plant having a
XX modified trait, the method comprising selecting a polynucleotide that
XX encodes a polypeptide or an antisense nucleic acid, inserting the
XX polynucleotide or antisense nucleic acid into an expression vector,
XX introducing the vector into a plant or a cell of a plant to overexpress
XX the polypeptide or antisense nucleic acid, thereby producing a modified

XX (CROP-) CROPDESIGN NV.
XX De Veylder L, Boudolf VKCK, Torres Acosta JA, Inze D;
XX WPI: 2000-431601/37.
XX P-PSDB; AAY96818.
XX Nucleic acids encoding plant cell cycle interacting proteins, useful
XX for regulating plant growth and in recombinant DNA protocols
XX Claim 1; Page 131-132; 152pp; English.
XX The VbBP clone is very similar to the A. thaliana putative DNA binding
XX protein and also contains a lot of homologues with Oryza sativa PCF2.
XX VbBP interacts with cyclin-dependent protein kinase (CDK) CDC2b but not
XX with CDC2a.
XX CDC2a and CDC2b are the only CDK genes to have been characterized in
XX detail in Arabidopsis thaliana. They were used as bait in a two-hybrid
XX screening assay with a cDNA library of a plant cell suspension as prey.
XX The plant cell cycle interacting proteins identified were designated
XX LRV15, PHO80-like protein, VB33, Vb89, VbDAP and VbHSP. The nucleic
XX acids, vectors comprising them, the proteins they express, antibodies
XX that bind to them and/or inhibitors of their protein expression and/or
XX activity may be used for modulating the cell cycle in an animal or plant,
XX plant cell division and/or growth, for influencing the activity of cell
XX cycle proteins in a plant or animal cell, as positive or negative
XX regulators of cell proliferation, for modifying the growth inhibition
XX caused by environmental stress conditions (e.g. to improve growth of
XX plants in normal or suboptimal nutrient conditions, especially
XX phosphorus), for use in a screening method for inhibitors or activators
XX of cell cycle protein, as growth regulators, herbicides and/or for
XX inducing nematode resistance in plants. The DNA sequences and their
XX regulatory sequences may be used as markers in plant or animal cell and
XX tissue cultures or as a marker in marker-assisted plant breeding. The
XX regulatory sequences may also be used for the expression of heterologous
XX DNA sequences during a stage of the cell cycle.
XX Q Sequence 1081 BP; 302 A; 195 C; 266 G; 318 T; 0 other;
Query Match 11.0%; Score 118; DB 21; Length 1081;
Best Local Similarity 73.3%; Pred. No. 3.4e-27;
Matches 151; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
211 AAGAGAGCTTCGACTAAACACCGTCACACGAGGTTGAGGAGGAGGAGGATCGG 270
232 AAGAGAGCTTCGACTAAACACCGTCACACGAGGTTGAGGAGGAGGAGGATCGG 291
271 ATGCTGTCACCGTGTGCGCTAGGATTTTCAATTAATTCGAGAGTTAGGTCACAAATCC 330
292 ATGCTGTCCTTTGTGCTGCTAGGATTTTCAATTCGAGAGTTAGGTCACAAATCT 351
331 GACGGGAAACGATTCGTTGGTGTGGAGACGCTGAGCGCGGATTAAGCCGCCAGC 390
352 GATGGTGAACCTATCCAGTGGCTGCTTCAACAGCTGAGCCATCGATTATTCAGCTACT 411
391 GGTACGGGAAACGTTCCGCCCATCGC 416
412 GGTTCAGGAACCTATACCGGCTCTGC 437
RESULT 9
BX62217
ID ABX62217 standard; DNA; 422 BP.
XX ABX62217;
XX ABX62217;
XX 25-FEB-2003 (first entry)
XX Arabidopsis thaliana expressed sequence related polynucleotide #332.
XX Transgenic plant; plant; genetically modified cell;
XX environmental stress; ribozyme creation; disease resistance;

stress tolerance; fungicide screening; insecticide screening;
gene; ds.

Gossypium hirsutum.

US2002040490-A1.

04-APR-2002.

26-JAN-2001; 2001US-0770423.

27-JAN-2000; 2000US-178512P.

(GORL/) GORLACH J.

(ANY/) AN Y.

(HAMI/) HAMILTON C M.

(PRIC/) PRICE J L.

(RAIN/) RAINES T M.

(YUY/) YU Y.

(RAME/) RAMEAKA J G.

(PAGE/) PAGE A.

(MATH/) MATHW A V.

(LEDF/) LEDFORD B L.

(WOES/) WOESSNER J P.

(HAAS/) HAAS W D.

(GARC/) GARCIA C A.

(KRIC/) KRICKER M.

(SLAT/) SLATER T.

(DAVI/) DAVIS K R.

(ALLE/) ALLEN K.

(HOFF/) HOFFMAN N.

(HURB/) HURBAN P.

Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;

Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;

Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;

Hurban P;

WPI; 2003-110411/10.

Novel Arabidopsis thaliana nucleic acid useful for identifying
homologous or related genes, and to create genetically modified and
transgenic organisms, such as plant cells and plants.

Claim 1; SEQ ID NO 332; 43pp; English.

The invention describes an Arabidopsis thaliana nucleic acid (I). The
polypeptide (II) encoded by (I), transgenic plant (III) or genetically
modified cell (IV) are useful for screening a candidate agent for its
biological effect, by combining the candidate agent with (II), (III) or
(IV), and determining the effect of the candidate agent on (II), (III) or
(IV). (I) is useful for identifying homologous or related genes, for
producing compositions that modulate the expression or function of its
encoded protein, for mapping functional regions of the protein, in
diagnosis, for studying associated physiological pathways, for genetic
manipulation of cells, preferably plant cells, in screening assays of
various plant strains to determine the strains that are capable of
withstanding a particular disease or environmental stress, for enhancing
or inhibiting production of bioactive product in a plant, for
producing polypeptides, as probes for the detection of mRNA in biological
samples, to generate additional copies of (I), to generate ribozymes or
oligonucleotides, as single stranded DNA probes or as triple-strand
forming oligonucleotides, and to create genetically modified and
transgenic organisms, such as plant cells and plants. (II) or (III) is
useful for introducing or improving disease resistance and stress
tolerance in plants, screening biological active agents, e.g., al pathways.
fungicides, insecticides, etc., and for elucidating biochemical resistance, enhanced
traits of interest, for screening programs, as crops which exhibit
enhanced tolerance to environmental stress, or to produce a factor. This
sequence represents a nucleic acid that may correspond to naturally
occurring Arabidopsis thaliana expressed sequences.
Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the
CC US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=999909770423.
XX
SQ Sequence 422 BP; 111 A; 100 C; 118 G; 93 T; 0 other;

Query Match 10.4%; Score 111.2; DB 25; Length 422;
Best Local Similarity 71.6%; Pred. No. 3.3e-25;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 209 TGRAGAGAGCTTCGACTAAAGACCGTCACAGAGGTTGAGGAGGAGGAGGAGTAC 268
Db 57 TTAAGAAACCAACCTCTAAAGATCGACACAGCAAGTTGACGAGGAGGAGGAGGATTC 126

QY 269 GGATGCTCCGACAGTGTGCGGTAGGATTTTCAATTAACTCGAGAGTTAGGTCACAAAAT 328
Db 127 GTATGCCAATCATTTGGCAGCTCGAGTTTTCATTTGACACAGAGTTAGGTCACAAGT 186

QY 329 CCGACGGGGAACCAATTCGGTGTGTTGTGGAGACGCTGAGCCGGCGGATTTATAGCGGCCA 388
Db 187 CCGATGTGTCAAACCATAGAGTGGGCTTCCGTCAGCTGAGCCTTCTATCATAGCGGCCA 246

QY 389 CCGGTACGGGAACGGTTCGCGCCA 412
Db 247 CTGGAATGGCACTACTCCGGCGA 270

RESULT 10
ABK65215
ID ABK65215 standard; cDNA; 1009 BP.
AC
XX
AC ABK65215;
XX
DT 02-JUL-2002 (first entry)
XX
DE Arabidopsis cDNA encoding a transcription factor #67.
XX
KW Plant; ss; gene; transcription factor; transgenic;
KW agriculture; metabolic chemical; environmental stress; drought;
KW microbial disease resistance; herbicide resistance; seed yield;
KW fruit yield; growth rate; leaf senescence; flower senescence.
XX
OS Arabidopsis thaliana.
XX
OS WO200215675-A1.
XX
PN 28-FEB-2002.
XX
PF 22-AUG-2001; 2001WO-US26189.
XX
PR 22-AUG-2000; 2000US-227439P.
PR 16-NOV-2000; 2000US-0713994.
PR 16-APR-2001; 2001US-0837944.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
FA
FA (PILG/) PILGRIM M.
FA (CREE/) CREELMAN R.
FA (DUBE/) DUBELL A. J.
FA (HEAR/) HEARD J.
FA (JIAN/) JIANG C.
FA (KEDD/) KEDDIE J.
FA (ADAM/) ADAM L.
FA (RATC/) RATCLIFF O.
FA (REUB/) REUBER J. L.
FA (RIEC/) RIECHMANN J. L.
FA (YUGG/) YU G.
FA (PINE/) PINEDA O.
XX
XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddle J;
PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX
XX WPI; 2002-292022/33.
DR P-PSDB; AAU93029.

XX An isolated or recombinant polynucleotide used to produce a transgenic
PT plant -
XX
XX Claim 4; Page 309-310; 941pp; English.
XX
CC The invention relates to 1 of 232 isolated or recombinant polynucleotides
CC encoding an Arabidopsis thaliana transcription factor, their variants,
CC complements, fragments, or related polynucleotide with 31% to 95%
CC sequence identity, where the plant possesses an altered trait as compared
CC to a wild-type or reference plant, or the plant exhibits an altered
CC phenotype as compared to a wild-type or reference plant, or the plant
CC exhibits ectopic expression or altered expression of one or more genes
CC associated with a plant trait as compared to a wild plant. Also included
CC are a transgenic plant comprising the polynucleotides, a computer
CC readable medium having stored sequence information, and identifying a
CC homologue sequence from a database comprising a plurality of known plant
CC sequences comprising inputting sequence information selected from one of
CC 464 fully defined sequences given in the specification. The isolated or
CC recombinant polynucleotide is used for producing a plant having a
CC modified trait, the method comprising selecting a polynucleotide that
CC encodes a polypeptide or an antisense nucleic acid, inserting the
CC polynucleotide or antisense nucleic acid into an expression vector,
CC introducing the vector into a plant or a cell of a plant to overexpress
CC the polypeptide or antisense nucleic acid, thereby producing a modified
CC plant, and selecting for a modified trait (e.g. increased
CC production of agriculturally useful proteins or metabolic chemicals,
CC pest tolerance, environmental stress response (e.g. drought), microbial
CC disease resistance, herbicide resistance, seed and fruit yield, growth
CC rate, leaf and flower senescence and many other traits listed in the
CC specification). The present sequence is one of the 232 polynucleotides
CC encoding an A. thaliana transcription factor.
XX
SQ Sequence 1009 BP; 256 A; 212 C; 267 G; 274 T; 0 other;

Query Match 10.4%; Score 111.2; DB 24; Length 1009;
Best Local Similarity 71.6%; Pred. No. 5.2e-25;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 209 TGRAGAGAGCTTCGACTAAAGACCGTCACAGAGGTTGAGGAGGAGGAGGAGTAC 268
Db 103 TTAAGAAACCAACCTCTAAAGATCGACACAGCAAGTTGACGAGGAGGAGGAGGATTC 162

QY 269 GGATGCTCCGACAGTGTGCGGTAGGATTTTCAATTAACTCGAGAGTTAGGTCACAAAAT 328
Db 163 GTATGCCAATCATTTGGCAGCTCGAGTTTTCATTTGACACAGAGTTAGGTCACAAGT 222

QY 329 CCGACGGGGAACCAATTCGGTGTGTTGTGGAGACGCTGAGCCGGCGGATTTATAGCGGCCA 388
Db 223 CCGATGTGTCAAACCATAGAGTGGGCTTCCGTCAGCTGAGCCTTCTATCATAGCGGCCA 282

QY 389 CCGGTACGGGAACGGTTCGCGCCA 412
Db 283 CTGGAATGGCACTACTCCGGCGA 306

RESULT 11
AAC33300
ID AAC33300 standard; DNA; 1486 BP.
XX
XX AAC33300;
AC
XX
AC AAC33300;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 4718.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.

XX	06-SEP-2000.	99US-0142055	PR	02-JUL-1999;
PD		99US-0142390.	PR	06-JUL-1999;
XX		99US-0142803.	PR	08-JUL-1999;
PF	25-FEB-2000; 2000EP-0301439.	99US-0142920.	PR	09-JUL-1999;
XX		99US-0142977.	PR	12-JUL-1999;
XX		99US-0143543.	PR	13-JUL-1999;
PR		99US-0143624.	PR	14-JUL-1999;
PR		99US-0144005.	PR	15-JUL-1999;
PR		99US-0144085.	PR	16-JUL-1999;
PR		99US-0144086.	PR	16-JUL-1999;
PR		99US-0144325.	PR	19-JUL-1999;
PR		99US-0144331.	PR	19-JUL-1999;
PR		99US-0144332.	PR	19-JUL-1999;
PR		99US-0144333.	PR	19-JUL-1999;
PR		99US-0144334.	PR	19-JUL-1999;
PR		99US-0144335.	PR	19-JUL-1999;
PR		99US-0144382.	PR	20-JUL-1999;
PR		99US-0144632.	PR	20-JUL-1999;
PR		99US-0144884.	PR	20-JUL-1999;
PR		99US-0144884.	PR	21-JUL-1999;
PR		99US-0145086.	PR	21-JUL-1999;
PR		99US-0145088.	PR	21-JUL-1999;
PR		99US-0145087.	PR	22-JUL-1999;
PR		99US-0145089.	PR	22-JUL-1999;
PR		99US-0145192.	PR	22-JUL-1999;
PR		99US-0145145.	PR	23-JUL-1999;
PR		99US-0145218.	PR	23-JUL-1999;
PR		99US-0145224.	PR	23-JUL-1999;
PR		99US-0145276.	PR	26-JUL-1999;
PR		99US-0145913.	PR	27-JUL-1999;
PR		99US-0145918.	PR	27-JUL-1999;
PR		99US-0145919.	PR	27-JUL-1999;
PR		99US-0145951.	PR	28-JUL-1999;
PR		99US-0146386.	PR	02-AUG-1999;
PR		99US-0146388.	PR	02-AUG-1999;
PR		99US-0146389.	PR	02-AUG-1999;
PR		99US-0147038.	PR	03-AUG-1999;
PR		99US-0147204.	PR	04-AUG-1999;
PR		99US-0147302.	PR	04-AUG-1999;
PR		99US-0147192.	PR	05-AUG-1999;
PR		99US-0147260.	PR	05-AUG-1999;
PR		99US-0147303.	PR	06-AUG-1999;
PR		99US-0147416.	PR	06-AUG-1999;
PR		99US-0147493.	PR	09-AUG-1999;
PR		99US-0147935.	PR	09-AUG-1999;
PR		99US-0148171.	PR	10-AUG-1999;
PR		99US-0148319.	PR	11-AUG-1999;
PR		99US-0148341.	PR	12-AUG-1999;
PR		99US-0148565.	PR	13-AUG-1999;
PR		99US-0148684.	PR	13-AUG-1999;
PR		99US-0149368.	PR	16-AUG-1999;
PR		99US-0149175.	PR	17-AUG-1999;
PR		99US-0149426.	PR	18-AUG-1999;
PR		99US-0149722.	PR	20-AUG-1999;
PR		99US-0149723.	PR	20-AUG-1999;
PR		99US-0149929.	PR	20-AUG-1999;
PR		99US-0149930.	PR	23-AUG-1999;
PR		99US-0149930.	PR	23-AUG-1999;
PR		99US-0150566.	PR	25-AUG-1999;
PR		99US-0150884.	PR	26-AUG-1999;
PR		99US-0151065.	PR	27-AUG-1999;
PR		99US-0151066.	PR	27-AUG-1999;
PR		99US-0151080.	PR	27-AUG-1999;
PR		99US-0151303.	PR	30-AUG-1999;
PR		99US-0151438.	PR	31-AUG-1999;
PR		99US-0151930.	PR	01-SEP-1999;
PR		99US-0152363.	PR	07-SEP-1999;
PR		99US-0153070.	PR	10-SEP-1999;
PR		99US-0153758.	PR	13-SEP-1999;
PR		99US-0154018.	PR	15-SEP-1999;
PR		99US-0154039.	PR	16-SEP-1999;
PR		99US-0154779.	PR	20-SEP-1999;

KW	microbial disease resistance; herbicide resistance; seed yield; fruit yield; growth rate; leaf senescence; flower senescence.
XX	Arabidopsis thaliana.
OS	WO200215675-A1.
PN	28-FEB-2002.
XX	22-AUG-2001; 2001WO-US26189.
PD	22-AUG-2000; 2000US-227439P.
XX	16-NOV-2000; 2000US-071394.
PF	16-APR-2001; 2001US-0837944.
PR	(MEND-) MENDEL BIOTECHNOLOGY INC.
XX	(PIUG/) PILGRIM M.
PA	(CREE/) CREELMAN R.
PA	(DUBE/) DUBELL A J.
PA	(HEAR/) HEARD J.
PA	(JIAN/) JIANG C.
PA	(KEDD/) KEDDIE J.
PA	(ADAM/) ADAM L.
PA	(RATC/) RATCLIFF O.
PA	(REUB/) REUBER J L.
PA	(RIEC/) RIECHMANN J L.
PA	(YUGG/) YU G.
PA	(PINE/) PINEDA O.
XX	Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J, Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
PI	WPI: 2002-292022/33.
DR	P-PSDB; AAU93167.
XX	An isolated or recombinant polynucleotide used to produce a transgenic plant -
PT	Claim 4; Page 845-846; 941pp; English.
XX	The invention relates to 1 of 232 isolated or recombinant polynucleotides encoding an Arabidopsis thaliana transcription factor, their variants, complements, fragments, or related polynucleotide with 31% to 95% sequence identity, where the plant possesses an altered trait as compared to a wild-type or reference plant, or the plant exhibits an altered phenotype as compared to a wild-type or reference plant, or the plant exhibits ectopic expression or altered expression of one or more genes associated with a plant trait as compared to a wild plant. Also included are a transgenic plant comprising the polynucleotides, a computer readable medium having stored sequence information, and identifying a homologue sequence from a database comprising a plurality of known plant sequences comprising inputting sequence information selected from one of 464 fully defined sequences given in the specification. The isolated or recombinant polynucleotide is used for producing a plant having a modified trait, the method comprising selecting a polynucleotide that encodes a polypeptide or an antisense nucleic acid, inserting the polynucleotide or antisense nucleic acid into an expression vector, introducing the vector into a plant or a cell of a plant to overexpress the polypeptide or antisense nucleic acid, thereby producing a modified plant, and selecting for a modified trait (e.g. increased production of agriculturally useful proteins or metabolic chemicals, pest tolerance, environmental stress response (e.g. drought), microbial disease resistance, herbicide resistance, seed and fruit yield, growth rate, leaf and flower senescence and many other traits listed in the specification). The present sequence is one of the 232 polynucleotides encoding an A. thaliana transcription factor.
CC	Sequence 1172 BP; 331 A; 217 C; 251 G; 373 T; 0 other;
XX	Query Match
XX	Best Local Similarity 9.3%; Score 100; DB 24; Length 1172;
XX	Matches 139; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

2y 209 TGAAGAGAGCTTCGACTAAGACCGGTCAACGAGAGTTGAAAGAGAGGAGAGATAC 268
2b 192 TGAAGAAACACCGGCGAAGATCGACATAGCAAGTGCATGGAAGAGGAGAGATCC 251
2y 269 GGATGCTCCCGCTGCGGTAGGATTTTCAATTAACCTGAGAGTTAGTCAAAAT 328
2b 252 GATGCGGATTATGTGCTGCTGCTGTTTTCAGCTAACGAGAGCTTGTCTAAGT 311
2y 329 CCGACGGCGAAGAGATTGCTGCTGTTTTCGAGAACGCTGAGCGCGATTATAGCGGCA 388
2b 312 CAGATGCCCAACAAATGAATGTTACTTGGTCAAGCAGAGCTTCTATTATAGTGCAA 371
2y 389 CGGTACGGGAGCGTTCCGGCA 412
2b 372 CAGGAATGGTCAACTCCAGCA 395

RESULT 13

ABK65216

ID ABK65216 standard; cDNA; 1263 BP.

AC

AC

AC

AC

02-JUL-2002 (first entry)

Arabidopsis cDNA encoding a transcription factor #68.

DE Plant; ss; gene; transcription factor; transgenic;
CW agriculture; metabolic chemical; environmental stress; drought;
CW microbial disease resistance; herbicide resistance; seed yield;
CW fruit yield; growth rate; leaf senescence; flower senescence.

Arabidopsis thaliana.

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

X

CC to a wild-type or reference plant, or the plant exhibits an altered
CC phenotype as compared to a wild-type or reference plant, or the plant
CC exhibits ectopic expression or altered expression of one or more genes
CC associated with a plant trait as compared to a wild plant. Also included
CC are a transgenic plant comprising the polynucleotides, a computer
CC readable medium having stored sequence information, and identifying a
CC homologue sequence from a database comprising a plurality of known plant
CC sequences comprising inputting sequence information selected from one of
CC 464 fully defined sequences given in the specification. The isolated or
CC recombinant polynucleotide is used for producing a plant having a
CC modified trait, the method comprising selecting a polynucleotide that
CC encodes a polypeptide or an antisense nucleic acid, inserting the
CC polynucleotide or antisense nucleic acid into an expression vector,
CC introducing the vector into a plant or a cell of a plant to overexpress
CC the polypeptide or antisense nucleic acid, thereby producing a modified
CC plant, and selecting for a modified trait (e.g. increased
CC production of agriculturally useful proteins or metabolic chemicals,
CC pest tolerance, environmental stress response (e.g. drought), microbial
CC disease resistance, herbicide resistance, seed and fruit yield, growth
CC rate, leaf and flower senescence and many other traits listed in the
CC specification). The present sequence is one of the 232 polynucleotides
CC encoding an A. thaliana transcription factor.

XX Sequence 1263 BP; 339 A; 305 C; 280 G; 339 T; 0 other;

SQ

Query Match 9.2%; Score 98.6; DB 24; Length 1263;

Best Local Similarity 55.4%; Pred. No. 6.9e-21;

Matches 191; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

Qy 68 CCATCATCAACGGCGTCAGAAACGTCGAAACCTTCAAGACCTTCCAAAGTAATCCACAG 127

Db 121 CCACACAAACACACAGAGCAACACACACACTGTTCCGCCCATCTGGTCCCATCCA 180

Qy 128 TGAGTCTCGAGCCCAAGCGGAGCCGCTGATGCGCTCGTTTCAATGCTTTAGTCCAC 187

Db 181 TGGGACCAATCTCGGTTCAGTCTCATTACACCACTGCTCAAACTCCATCCACCA 240

Qy 188 CGTCTTCGACAGGACCAACCATTCGAGAGAGCTTCGACTAAAGACCGTCACAGAGTTG 247

Db 241 CCGTCACCGCGCTAAACACACCGCAACCGCTCAAGGACCGTCACATCAAGTAG 300

Qy 248 AAGGAGAGGAGAGAGATACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307

Db 301 ACGGACGTGCGCGGAGATACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360

Qy 308 CTCGAGAGTTAGTTCACAAATCCGAGCGCAACAGATTGCTGCTGCTGCTGCTGCTGCTG 367

Db 361 CACGTGAGTTACACCAATCCGAGCGGAGACTATAGATGCTGCTGCTGCTGCTGCTGCTG 420

Qy 368 AGCGGCGATTATAGCGCGCAACCGGTACCGGAAACGGTTCCCGCA 412

Db 421 AGCCAGCTATCATCGCAGCCAGCAACCGGAACTGGAAACCATACCGGCA 465

RESULT 14

AAC44188

ID AAC44188 standard; DNA; 1379 BP.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Arabidopsis thaliana DNA fragment SEQ ID NO: 41952.

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

An isolated or recombinant polynucleotide used to produce a transgenic
plant -

Claim 4; Page 312-314; 941pp; English.

The invention relates to 1 of 232 isolated or recombinant polynucleotides
encoding an Arabidopsis thaliana transcription factor, their variants,
complements, fragments, or related polynucleotide with 31% to 95%
sequence identity, where the plant possesses an altered trait as compared

XX PR 25-FEB-2000; 2000EP-0301439. 99US-0121825.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131149.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139750.
XX PR 18-JUN-1999; 99US-0139763.
XX PR 21-JUN-1999; 99US-0139817.
XX PR 22-JUN-1999; 99US-0139899.
XX PR 23-JUN-1999; 99US-0140353.
XX PR 23-JUN-1999; 99US-0140354.
XX PR 23-JUN-1999; 99US-0140695.
XX PR 28-JUN-1999; 99US-0140823.
XX PR 29-JUN-1999; 99US-0140991.
XX PR 30-JUN-1999; 99US-0141287.
XX PR 01-JUL-1999; 99US-0141842.
XX PR 01-JUL-1999; 99US-0142154.
XX PR 02-JUL-1999; 99US-0143055.
XX PR 06-JUL-1999; 99US-0143390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

24-SEP-1999; 99US-0155659.
28-SEP-1999; 99US-0156458.
29-SEP-1999; 99US-0156596.
04-OCT-1999; 99US-0157117.
05-OCT-1999; 99US-0157753.
06-OCT-1999; 99US-0157865.
07-OCT-1999; 99US-0158029.
08-OCT-1999; 99US-0158232.
12-OCT-1999; 99US-0158369.
13-OCT-1999; 99US-0159293.
13-OCT-1999; 99US-0159294.
13-OCT-1999; 99US-0159295.
14-OCT-1999; 99US-0159329.
14-OCT-1999; 99US-0159330.
14-OCT-1999; 99US-0159331.
14-OCT-1999; 99US-0159637.
14-OCT-1999; 99US-0160814.
21-OCT-1999; 99US-0160815.
22-OCT-1999; 99US-0160980.
22-OCT-1999; 99US-0160981.
22-OCT-1999; 99US-0160989.
25-OCT-1999; 99US-0161404.
25-OCT-1999; 99US-0161405.
25-OCT-1999; 99US-0161406.
26-OCT-1999; 99US-0161359.
26-OCT-1999; 99US-0161360.
26-OCT-1999; 99US-0161361.
28-OCT-1999; 99US-0161920.
28-OCT-1999; 99US-0161992.
28-OCT-1999; 99US-0161993.
29-OCT-1999; 99US-0162142.

Query Match 8.7%; Score 93.2; DB 21; Length 1379;
Best Local Similarity 68.8%; Pred. No. 48-19; 58; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
209 TGAAGAGAGCTTCGACTAAGACCGTCAACGAGGTTGAAGGAGAGGAGGAGGATAC 268
195 TGAAGAAACACCGCGGAAGATCGACATAGCAAGTGTGATGGAAGAGGAGGAGGATCC 254
269 GATGCTCCGACGCTGCGGCTAGGATTTTCAATTAATCTGAGAGCTAGTCAAAAT 328
255 GTATGCCGATTATATGTGCTGCTGTTTTCAGCTAACGAGAGAGCTTGTCTAAT 314
329 CGACGGCGAAACGATTCGTTGTTGAGAACGCTGAGCGCGGATTATAGCGGCA 388
315 CAGATGCCAAACAAATGAATGTTACTTCGTCACGAGAGGCTTCTATTATAGTGCAA 374
389 CGGGTA 394
375 CAGGGA 380

RESULT 15
ABN85767
ID ABN85767 standard; cDNA; 83698 BP.

AC ABN85767;

21-OCT-2002 (first entry)

Arabidopsis yellow stripe-like 4 encoding cDNA SEQ ID NO 9.

Maize; transgenic; plant; yellow stripe-like; ysl; Arabidopsis;
iron uptake, bioremediation; yellow stripe 1; ysl; gene; ss.

Arabidopsis sp.

XX WO200240688-A2.
XX 23-MAY-2002.
XX 16-NOV-2001; 2001WO-US43101.
XX 16-NOV-2000; 2000US-249222P.
XX (UYUA) UNIV YALE.
XX Walker EL, Dellaporta S;
XX WPI; 2002-490144/52.
XX P-PSDB; AB83920.
XX New yellow stripe and yellow stripe-like genes, useful for altering
XX the distribution of iron within the plant body so that edible parts of
XX crop plants have more iron, or for producing plants useful in enhancing
XX iron uptake from soil -
XX Claim 1; Page 114-159; 187pp; English.
XX The invention relates to an isolated nucleic acid molecule (I), maize
XX yellow stripe 1 (ysl) or yellow stripe-like (ysl) from Arabidopsis
XX (ABN85763-ABN85771). (I) is useful for generating transgenic plants which
XX can be used for enhancing iron uptake from soil and for bioremediation of
XX metal or heavy metal contaminated soil. (I) may also be used to alter the
XX distribution of iron within the plant body so that edible parts of crop
XX plants have more iron. Transgenic plants may also be used in conventional
XX plant breeding schemes to produce progeny which also contain the gene of
XX interest. The present sequence is that of the Arabidopsis ysl encoding
XX cDNA of the invention.
XX SQ Sequence 83698 BP; 26295 A; 15150 C; 15104 G; 27149 T; 0 other;

Query Match 8.3%; Score 89; DB 24; Length 83698;
Best Local Similarity 65.2%; Pred. No. 8.3e-17;
Matches 131; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
Qy 210 GAAGAGAGCTTCGACTAAGACCGTCAACGAGGTTGAAGGAGGAGGAGGATACG 269
Db 18681 GAAGAAGAACCAACAAAGATCGTCACTTAAGTTGAAGGAGGAGGATGAGTTAG 18740
Qy 270 GATGCTCCGACCGTGTGCGGCTAGGATTTTCAATTAATCTGAGAGTTAGTCAAAATC 329
Db 18741 GTTACCTCCACTCTGTGACGAGGATTTATCAATTGACTAAAGATTAGTCAAAATC 18800
Qy 330 CGACGGCGAAACGATTCGTTGTTGAGAACGCTGAGCGCGGATTATAGCGGCA 389
Db 18801 AGATGTTGAGACTCTTGAATGTTGCTTCAACATGCTGAGCGCATCTCTGCTAC 18860
Qy 390 GGGTACGGGAACGGTTCCCGC 410
Db 18861 TGTAAATGTTATCAAAACCCAC 18881

Search completed: February 1, 2004, 21:58:51
Job time : 322.739 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:45:16 / Search time 70.5426 Seconds
(without alignments)
6701.220 Million cell updates/sec

Title: US-09-938-842A-1034

Perfect score: 1071

Sequence: 1 atggcgacattcgaagct.....cagcgatcgcgaaccactga 1071

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	44	4.1	7218	1	US-08-232-463-14
2	41	3.8	7218	1	US-08-232-463-14
3	33	3.1	2127	4	US-09-252-991A-8192
4	33	3.1	2874	4	US-09-252-991A-8112
5	32.8	3.1	4403765	3	US-09-103-840A-2
6	32.2	3.0	2406	4	US-08-632-098-5
7	32.2	3.0	2439	4	US-08-632-098-5
8	31.4	2.9	4411529	3	US-09-103-840A-1
9	30.6	2.9	412	3	US-08-961-083-111
10	30.6	2.9	412	4	US-08-961-083-111
11	30.6	2.9	1288	4	US-09-536-784-111
12	30.6	2.9	6693	4	US-08-620-312D-546
13	30.4	2.8	2172	1	US-07-982-712-1
14	30.4	2.8	7766	4	US-08-135-619-3
15	30.4	2.8	580073	4	US-08-585-828D-1
16	30.2	2.8	801	3	US-08-598-416-436
17	30.2	2.8	1553	3	US-09-217-490-1
18	30.2	2.8	2396	4	US-09-221-017B-74
19	30.2	2.8	23673	4	US-09-773-816-1
20	30	2.8	536	3	US-08-714-918-57
21	30	2.8	536	3	US-09-265-315-57
22	30	2.8	536	3	US-09-265-315-57
23	30	2.8	536	3	US-09-266-417-57
24	30	2.8	1664976	4	US-08-936-421B-1
25	29.8	2.8	582	4	US-09-252-991A-12349
26	29.8	2.8	618	4	US-09-252-991A-12199
27	29.8	2.8	1026	4	US-09-252-991A-12295

28 29.8 2.8 2628 4 US-09-294-531B-5
29 29.8 2.8 4062 4 US-09-520-312B-348
30 29.8 2.8 8878 1 US-08-759-444-2
31 29.8 2.8 9880 3 US-08-680-897-1
32 29.8 2.8 38155 4 US-09-453-702B-79
33 29.8 2.8 4403765 3 US-09-103-840A-2
34 29.8 2.8 4411529 3 US-09-103-840A-1
35 29.8 2.8 1302 2 US-08-529-600B-3
36 29.6 2.8 1302 2 US-08-973-275-6
37 29.6 2.8 1302 2 US-09-122-632-3
38 29.6 2.8 4065 4 US-09-016-434-1105
39 29.6 2.8 4739 3 US-08-585-871-1
40 29.6 2.8 33960 4 US-09-453-702B-11
41 29.4 2.7 278 4 US-09-313-294A-5277
42 29.4 2.7 289 3 US-09-007-005-17
43 29.4 2.7 289 3 US-09-244-796-17
44 29.4 2.7 2277 1 US-08-676-967-2
45 29.4 2.7 2277 1 US-08-676-974-2

ALIGNMENTS

RESULT 1

US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:

APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463

FILING DATE:

PRIOR APPLICATION DATA:

CLASSIFICATION: 435

PRIOR APPLICATION NUMBER: US/07/935,313

FILING DATE: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMD

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: pTZsept-Fls

US-08-232-463-14

Query Match 4.1%; Score 44; DB 1; Length 7218;

Db 130 AATACCGGACCCGATCGCGGATACCAAGAACACCCCTGAAGCCGCGGAGCGGACCG 189
Qy 231 CCCTCACACGAAGTTGAAGAAAGAGGAGGATACCGATCCCTGCCACGCTGTGC 287
Db 190 TCGTGTCTGAAGACTTCATCATGCGGAGAGATCAACCCACTTCGACCAACGAGCGC 246

RESULT 4

US-09-252-991A-8112/c
; Sequence 8112, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8112
; LENGTH: 2874
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (2266)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-8112

Query Match 3.1%; Score 33; DB 4; Length 2874;
Best Local Similarity 49.2%; Pred. No. 2.5;
Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
Qy 111 CCAAGTAAATCCACAGTAGTGTCTGAGCCCAAGGCGGAGCGGTGATCGCTGCTTTC 170
Db 2487 CCAAGCTGGACAGCTGGAGGCTATAGCAGACGACGCCACCGGAGGCGCTGAGCAC 2428
Qy 171 AATGCTTTAGTCCACCGTCTTCGACAGACACCACTTGAAGAGAGCTTCGACTAAAGA 230
Db 2427 AATACCGGACCCGATCGCGGATACCAAGAACACCCCTGAAGCCGCGGAGCGGACCG 2368
Qy 231 CCCTCACACGAAGTTGAAGAAAGAGGAGGATACCGATCCCTGCCACGCTGTGC 287
Db 2367 TCGTGTCTGAAGACTTCATCATGCGGAGAGATCAACCCACTTCGACCAACGAGCGC 2311

RESULT 5

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.1%; Score 32.8; DB 3; Length 4403765;
Best Local Similarity 59.8%; Pred. No. 68;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
Qy 133 CTCGAGCCCAAGCGGAGCGCGGTGATCGCTGCTTTTCAATGTCTTTAGCTCCACCGTCT 192
Db 3941480 CTGCGGACAAAGCGGCGCGGGGTGCTGCGGGGCGCGCGGATACCCACCGGCA 3941539
Qy 193 TCGACAGGACCCACCTTGAAGAGAGCTTCGAC 224
Db 3941540 TCGCGGACCCGCGGTGACGCGGCGCACCGC 3941571

RESULT 6

US-09-632-098-5
; Sequence 5, Application US/09632098
; Patent No. 6420154
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Baidur, Nand
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
; FILE REFERENCE: 99-39
; CURRENT APPLICATION NUMBER: US/09/632,098
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 2406
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(2406)
; OTHER INFORMATION: n = A, T, C or G
US-09-632-098-5

Query Match 3.0%; Score 32.2; DB 4; Length 2406;
Best Local Similarity 25.2%; Pred. No. 4.1;
Matches 82; Conservative 47; Mismatches 197; Indels 0; Gaps 0;
Qy 637 CTGCGGACGAATGATCCGATGTGGCTATTCCATCAAGCAATGATTCGAGCGTC 696
Db 1016 CNGCNGCNAATGGCNCAYGAATHGNCAYWSNTNGNNTNWSNCAYGAYCCNGAYG 1075
Qy 697 GGAGCTTTCTTTGATTCCACAAATCGCTGCTCCGATCGAATCAGCCTCAGTTATTAGCT 756
Db 1076 GNTGYGTGNGAGCNGCNGCNGARWSNGGNGTGYTNATGCGCNGCNGCNGCNGC 1135
Qy 757 TTTCCGCGCGCGCTGCTTCGCGCTGTCTTAGCTGCGCGCTGTTCAAGAGGCTCCAGC 816
Db 1136 AYCNTTYCCNMGNGNTTYWSNGCNTGYWSNMGNCARYTNMGNCNTTYTYMGNA 1195
Qy 817 ATGGCTAGACCACTCTTTTACAAGTTGTTCCAAAGCAGCGGCTTTGTATCCGTTTCAGAC 876
Db 1196 ARGNGGNGGNGCNTGYTNWSNAAVGCNCCNGAYCNGGNTNCCNGCNGCNY 1255
Qy 877 GTTAGCGTTTGAATTTATCAAGAGCGAGCGTGGTTATGGCTCCGAGCTCAAGCTCAGGC 936
Db 1256 TMTGYGNAAYGNTTYGTNGARGCNGGNGARGARTGYGAYTYGNGCNGCNGCARGART 1315
Qy 937 GTACACACCGGTAGTTCATCGTCAAT 962
Db 1316 GYMGNGAYTNTGYTGTTCNCAY 1341
RESULT 7
US-09-632-098-6
; Sequence 6, Application US/09632098


```

; Patent No. 6420154
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Baindur, Nand
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
; FILE REFERENCE: 99-39
; CURRENT APPLICATION NUMBER: US/09/632,098
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate sequence
; NAME/KEY: misc.feature
; LOCATION: (1)...(2439)
; OTHER INFORMATION: n = A,T,C or G
US-09-632-098-6

Query Match      3.0%; Score 32.2; DB 4; Length 2439;
Best Local Similarity 25.2%; Pred. No. 4.1;
Matches 82; Conservative 47; Mismatches 197; Indels 0; Gaps 0;

QY 637 CTGGCGCAGGAATGATCCGATGGGCTATTCATCAACGCAATGATTCGACGGTC 696
DB 1016 CNGCNGCAGCATGCGCAYGARATHGNCAYWNYNGNYTWSNCAYGAYCCNGAYG 1075

QY 697 GGAGCTTTCTTTGATTCACAAATCGCTCGTCGATCAAGCTCAGTTATTAGCT 756
DB 1076 GNTGYGTGNGGNGCNGCNGCNGGNGGNGTGTGATGCGNGCNGCNGCNGC 1135

QY 757 TTTCCCGCGCGCTGCTTCGCGCTGCTTACGTCGCGCTGTTCAACAGGCTTCAAG 816
DB 1136 AYCNTTYCCNMGNTNTYWSNGNTGYWSNMGNGNCARYTNMGNGCNTTYYTNGNA 1195

QY 817 ATGCTAGACCACTCCCTTTACAAGTTGTTCCAGACGCGCTTGTATCGTTTCAGAC 876
DB 1196 ARGNGGNGGNGCNTGYTWSNAYGNCNGCNGAYCCNGGNYTNCNGTNCNGCNGNY 1255

QY 877 GTTAGCGTTGCAATTTATCAAGAGCAGCTCGGTTATGCTCGAGCTCAAGCTCAGG 936
DB 1256 TMTGYGNAAYGNTTYGTNGARGCNGGNGARGARTGYGTYGNGCNGGNGCARGART 1315

QY 937 GTAACAACCGGTAGTTCATCGTCAAT 962
DB 1316 GYMNGAYTNTGYTTCNCAY 1341

RESULT 8
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

; Patent No. 6420154
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Baindur, Nand
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
; FILE REFERENCE: 99-39
; CURRENT APPLICATION NUMBER: US/09/632,098
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate sequence
; NAME/KEY: misc.feature
; LOCATION: (1)...(2439)
; OTHER INFORMATION: n = A,T,C or G
US-09-632-098-6

Query Match      3.0%; Score 32.2; DB 4; Length 2439;
Best Local Similarity 25.2%; Pred. No. 4.1;
Matches 82; Conservative 47; Mismatches 197; Indels 0; Gaps 0;

QY 637 CTGGCGCAGGAATGATCCGATGGGCTATTCATCAACGCAATGATTCGACGGTC 696
DB 1016 CNGCNGCAGCATGCGCAYGARATHGNCAYWNYNGNYTWSNCAYGAYCCNGAYG 1075

QY 697 GGAGCTTTCTTTGATTCACAAATCGCTCGTCGATCAAGCTCAGTTATTAGCT 756
DB 1076 GNTGYGTGNGGNGCNGCNGCNGGNGGNGTGTGATGCGNGCNGCNGCNGC 1135

QY 757 TTTCCCGCGCGCTGCTTCGCGCTGCTTACGTCGCGCTGTTCAACAGGCTTCAAG 816
DB 1136 AYCNTTYCCNMGNTNTYWSNGNTGYWSNMGNGNCARYTNMGNGCNTTYYTNGNA 1195

QY 817 ATGCTAGACCACTCCCTTTACAAGTTGTTCCAGACGCGCTTGTATCGTTTCAGAC 876
DB 1196 ARGNGGNGGNGCNTGYTWSNAYGNCNGCNGAYCCNGGNYTNCNGTNCNGCNGNY 1255

QY 877 GTTAGCGTTGCAATTTATCAAGAGCAGCTCGGTTATGCTCGAGCTCAAGCTCAGG 936
DB 1256 TMTGYGNAAYGNTTYGTNGARGCNGGNGARGARTGYGTYGNGCNGGNGCARGART 1315

QY 937 GTAACAACCGGTAGTTCATCGTCAAT 962
DB 1316 GYMNGAYTNTGYTTCNCAY 1341

RESULT 8
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      2.9%; Score 31.4; DB 3; Length 4411529;
Best Local Similarity 59.6%; Pred. No. 95;
Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 136 GAGCCCAAGGCGGAGCGGTATGCGTCGTTTCAATGTTTGTAGTCCACCGCTCTCG 195
DB 3947108 GCGGACAAAGCGCGCGCGGGGTGCTGGCGGGCGCGCGGATACCCACCGGATCG 3947167

QY 196 ACAGGACCAACCATTCAGAGAGAGCTTCCGAC 224
DB 3947168 GCGGACCGCGGTGACGCGCGCACCGGC 3947196

RESULT 9
US-08-961-083-111
; Sequence 111, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-111

Query Match      2.9%; Score 30.6; DB 3; Length 412;
Best Local Similarity 53.8%; Pred. No. 5.5;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 798 TGTTCACAGCGTTCCAGATGGCTAGACCACTCTTTTACAGTTGTTCCAGACGCG 857
DB 177 TGTCCAATCGCTCTTTATCATGTTGCGACTGCGCCACAGACCCCTTTTTCAGAGGGGG 236

QY 858 CTTTGTATCCGTTTTCAGAGCTTAGCGGTTCGAATTTATCAAGAGCAGCTCGGTAT 914
DB 237 CTTTGTCTCAGTACCGCTCTTATCGCTGCGCGTGTGTGAGGAGTGCCTGCTTTAT 293

RESULT 10
US-09-536-784-111
; Sequence 111, Application US/09536784
; Patent No. 6573082
```

GENERAL INFORMATION:
APPLICANT: Choei et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 111:
JS-09-536-784-111

Query Match 2.9%; Score 30.6; DB 4; Length 412;
Best Local Similarity 53.8%; Pred. No. 5.5;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

2Y 798 TGTTCACAGGCTTCACAGCTGCTAGACACCTCTTTACAGTGTTCACAGCGG 857
Db 177 TGCCCAATCGCTTTATCATGTTGCGACTGCGTCCACAGACCCCTTTTCAAAGGGGG 236
2Y 858 CTGTGTATCGTTTCAGACGTTAGCGTTTCGAATTTATCAAGAGCGAGCTCGGTTAT 914
Db 237 CTGTGTACAGTACCGCTGTTATCGCTCGCGTGTGTCAGAGTGCCGTGCTTTAT 293

RESULT 11
JS-09-620-312D-546/c
Sequence 546, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunding
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast

GENERAL INFORMATION:
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt FL_genes Version 1.0
SEQ ID NO 546
LENGTH: 1288
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (95)...(928)
US-09-620-312D-546

Query Match 2.9%; Score 30.6; DB 4; Length 1288;
Best Local Similarity 55.0%; Pred. No. 9.9;
Matches 60; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 7 ACATTTCAGAGTTGAAGAAGTTGCAGGCAAGATCAAACTCTAAGAGCCGTTGATCTA 66
Db 439 ATAGTCGGACCTCCACCCTTTGCGCTCCAGATGTTTCAGAGTCTCTGTCATCAA 380
Qy 67 ACCATCATCAACGGCGTCAAGAACTTCAGAACTTCAGACCTTTCCAAAG 115
Db 379 ATCATTATCAAGGCGATCTGGATGGTCAAAATTTGACTGTCTTTCAG 331

RESULT 12
US-08-961-527-195/c
Sequence 195, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 6693 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
US-08-961-527-195

Query Match 2.9%; Score 30.6; DB 4; Length 6693;
Best Local Similarity 53.8%; Pred. No. 23;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 798 TGTTCAACAGGCTTCACAGATGCTAGACACCTCTTTTACAGTGTTCACAGCGG 857
DB 2197 TGCCCAATCGCTCTTTATCATGTGTGCACTGCTCCACAGACCCCTTTTTCAGAGGGGG 2138
QY 858 CTTTGTATCGGTTTCAGAGCTTACGGTTTCGAATTTATCAAGAGGAGCGTGGTTAT 914
DB 2137 CTTTGTCTAGTACCGCTGTATCGTCTGGGCTGTGTGTCAGAGTGCCTGTCTTTAT 2081

RESULT 13
US-07-982-712-1
; Sequence 1, Application US/07982712
; Patent No. 5436391
; GENERAL INFORMATION:
; APPLICANT: Hideya FUJIMOTO, Kimiko ITOH
; APPLICANT: Mikihiro YAMAMOTO, and KO SHIMAMOTO
; TITLE OF INVENTION: Insecticidal protein-encoding Gene, Gramineous
; TITLE OF INVENTION: Plants transformed with the Gene, and Production Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Penack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 144 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/982,712
FILING DATE: 19921127
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:

TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2172 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: peptide
LOCATION: 1 to 2172
IDENTIFICATION METHOD: E or S
OTHER INFORMATION:
US-07-982-712-1

Query Match 2.8%; Score 30.4; DB 1; Length 2172;
Best Local Similarity 54.5%; Pred. No. 15;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 880 ACGGTTCCGAATTTATCAAGACGCGTGTGCTTATGGCTCCGAGCTCAAGCTCAGGGGTA 939

DB 821 AGAATTTCCAGCGCAGCTTCGGGGCTCCGCCAGGCGCATCGAGGCGAGCATCAGGAGCC 880
QY 940 ACAACGGTAGTTTCATCGTCAATTGCAACAACAACGACGACGACGCTGAGAG 991
DB 881 CGACCTCATGGACATCTCTCAACAGCATCACCATTCTACCGACGCCACAG 932

RESULT 14
US-09-125-619-3/c
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: US/09/125,619
; CURRENT APPLICATION NUMBER: US/09/125,619
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match 2.8%; Score 30.4; DB 4; Length 7766;
Best Local Similarity 48.3%; Pred. No. 28;
Matches 85; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 542 CCCTTTTCAGCTTCCTCCGGTTTAGTCCCAATGCCACGACGACGATCCAACTCCGC 601
DB 3151 CCTATCAGCAGCCTTCGACGACGACGATATCCCAACTTCTCAATTGCAGAGTTC 3092
QY 602 AGCTCTGGCATCATCCACTGTGGCTCAGCAACTTTCGCGCAAGAAATGTATCGATGT 661
DB 3091 CCTATAGCCCCCTCAGCTTCCCTTCTCATCATTTTCACAGCAAACTTCCATCCT 3032
QY 662 GGCTATTTCATCAACGCAATGATTCGACGCTCGGAGCTTCTCTTGTATCCA 717
DB 3031 TAGCCATCCCTCAAGCAATAGCAGCAGCAATCTGATCATCTCTTCTCATCCA 2976

RESULT 15
US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment:
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium

US-08-545-528D-1

Query Match 2.8%; Score 30.4; DB 4; Length 580073;
 Best Local Similarity 47.0%; Pred. No. 2.1e+02;
 Matches 94; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY	871	TCAGAGTTAGCGGTTGGAATTTATCAGAGCGACGTCGGTTATGGCTCGAGCTCAAGC	930
DB	3402	TCACCTCTTTCCATTACAGAGAAATCAGGAACAACCTCAACAATGTTCCATGCTCTTTT	3343
QY	931	TCAGGCGTAAACACCGGTAGTTTCATCGTCAATTGCAACAACAACGACGACACGCTGAGA	990
DB	3342	TCAGAGTTACCTTCTTGGACCAATCTCCAAATTACCTTTCTCCATCGCTAAAGCTGAGA	3283
QY	991	GACTTCTCCCTAGAGATATACGAGAAACAGAGCTTCACCGGTTCTATGAGCACCACAACA	1050
DB	3282	AAATACCTTTTATTTTTCAGCAAAACCTCAAACTTTAAAGAGAACTAAAGCGCATTAAACA	3223
QY	1051	GCACGGTCAATCGAACCACTG	1070
DB	3222	ACTGATGCACCAACACCGTG	3203

Search completed: February 2, 2004, 03:00:11
 Job time : 101.543 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:42:55 ; Search time 402.599 Seconds
(without alignments)
9695.297 Million cell updates/sec

Title: US-09-938-842A-1034

Perfect score: 1071

Sequence: 1 atggcgacaattcagaagct.....caggtcaccgaaccactga 1071

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10D_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1071	100.0	1071	10	US-09-938-842A-1034
2	1071	100.0	1071	12	US-09-938-842A-1034
3	448	41.8	460	10	US-09-924-035A-502
4	439	41.0	453	9	US-09-770-444-615
5	185	17.3	185	9	US-09-770-698-257
6	185	12.0	128	11	US-09-934-455-169
7	129	12.0	128	13	US-10-225-068-165
8	124.6	11.6	1231	11	US-09-934-455-137
9	124.6	11.6	1231	12	US-10-302-267-61
10	124.6	11.6	1231	13	US-10-225-068-245
11	121.6	11.4	1604	15	US-10-295-403-147
12	111.2	11.0	1009	11	US-09-770-961-675
13	111.2	10.4	1009	11	US-09-934-455-133
14	111.2	10.4	1009	13	US-10-225-068-135
15	111.2	10.4	1009	15	US-10-295-403-145

16	111.2	10.4	1009	16	US-10-278-536-203	Sequence 203, App
17	104.6	9.8	819	12	US-10-260-238-383	Sequence 383, App
18	100	9.3	1172	11	US-09-934-455-461	Sequence 461, App
19	98.6	9.2	1263	11	US-09-934-455-135	Sequence 135, App
20	89	8.3	955	13	US-10-225-068-181	Sequence 181, App
21	79.6	7.4	419	10	US-09-878-574-4107	Sequence 4107, App
22	77	7.2	612	12	US-10-260-238-5623	Sequence 5623, App
23	67.6	6.3	552	12	US-10-260-238-65	Sequence 65, Appl
24	50	4.7	228	12	US-10-260-238-5848	Sequence 5848, App
25	49.6	4.6	297	12	US-10-260-238-5405	Sequence 5405, App
26	43.6	4.1	447	9	US-09-770-444-761	Sequence 761, App
27	36.6	3.4	3316	15	US-10-242-056-58	Sequence 58, Appl
28	34	3.2	654	8	US-08-781-986A-613	Sequence 613, App
29	33.4	3.1	629	13	US-10-027-633-221892	Sequence 221892, App
30	33.4	3.1	629	14	US-10-027-633-221892	Sequence 221892, App
31	33.2	3.1	469	9	US-09-864-761-4365	Sequence 4365, App
32	33.2	3.1	828	15	US-10-156-761-5087	Sequence 5087, App
33	33.2	3.1	1893	15	US-10-128-714-7597	Sequence 7597, App
34	33.2	3.1	9025608	15	US-10-156-761-1	Sequence 1, Appl
35	33	3.1	671	15	US-10-184-644-346	Sequence 346, App
36	33	3.1	1285	15	US-10-184-634-346	Sequence 346, App
37	32.8	3.0	656	13	US-10-198-846-6773	Sequence 6773, App
38	32.6	3.0	656	14	US-10-027-633-151527	Sequence 151527, App
39	32.6	3.0	846	12	US-10-320-797-2065	Sequence 2065, App
40	32.6	3.0	1019	12	US-10-320-797-1065	Sequence 1065, App
41	32.6	3.0	2320	15	US-10-268-441-5	Sequence 5, Appl
42	32.6	3.0	3019	12	US-10-320-797-65	Sequence 65, Appl
43	32.4	3.0	1200	10	US-09-887-576-731	Sequence 731, App
44	32.4	3.0	429	15	US-10-268-441-3	Sequence 3, Appl
45	32.2	3.0				

ALIGNMENTS

RESULT 1

US-09-938-842A-1034
Sequence 1034, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPT300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1034
LENGTH: 1071
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1034

Query Match 100.0%; Score 1071; DB 10; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCGACAAATTCAGAAGCTTCAGAGAGTTGAGCAGCAAGATCAACTCTAAGAGCGGTT 60

Db 1 ATGGCGACAAATTCAGAAGCTTCAGAGAGTTGAGCAGCAAGATCAACTCTAAGAGCGGTT 60

Qy 61 GATCTAACCATCATCAACGGCGTCGAAACGTCGAAACCTTCAAGACCTTCCAAAGTAAT 120

Db 61 GATCTAACCATCATCAACGGCGTCGAAACGTCGAAACCTTCCAAAGTAAT 120

121 CCCACAGTGAAGTCTCGAGCCCAAGCCGAGCCGGTATGCGGTGCTTTCAATGTCCTTTA 180
121 CCCACAGTGAAGTCTCGAGCCCAAGCCGAGCCGGTATGCGGTGCTTTCAATGTCCTTTA 180
181 GCTCCAGCTCTTCGACAGGACCCACCATGAGAGAGCTTCGACTAAGACCGTCACACG 240
181 GCTCCAGCTCTTCGACAGGACCCACCATGAGAGAGCTTCGACTAAGACCGTCACACG 240
241 AAGGTTGAAGGAGAGGAGAGATACGATGCTGCCACGTCGTCGGGTAGGATTTTT 300
241 AAGGTTGAAGGAGAGGAGAGATACGATGCTGCCACGTCGTCGGGTAGGATTTTT 300
301 CAATTAATCGAGATTTAGTCTCAAAATCCGACGCGCAACGATTCGTTGGAG 360
301 CAATTAATCGAGATTTAGTCTCAAAATCCGACGCGCAACGATTCGTTGGAG 360
361 AACGCTGAGCCGCGGATTTATAGCCGCGCACCGGATCGGAAACGCTTCGCCCATCGCCATG 420
361 AACGCTGAGCCGCGGATTTATAGCCGCGCACCGGATCGGAAACGCTTCGCCCATCGCCATG 420
421 TCGGTTAAGCGAACTTTAAATATCCGACGACGACGACGATCTGATTCGATGCGGTGAA 480
421 TCGGTTAAGCGAACTTTAAATATCCGACGACGACGACGATCTGATTCGATGCGGTGAA 480
481 AATCTGATGAGAGAGAGAGGATTAAGACGCTTCTAACAGTGAATATAGACATAGCGAC 540
481 AATCTGATGAGAGAGAGAGGATTAAGACGCTTCTAACAGTGAATATAGACATAGCGAC 540
541 GCCGTTTCAGCTTCTCGGTTTATAGTCTCAATTCGACGACGACGATCCCAACCTCCG 600
541 GCCGTTTCAGCTTCTCGGTTTATAGTCTCAATTCGACGACGACGATCCCAACCTCCG 600
601 CAAGCTCTGGCATCATCCACTGCTGGCTCAGCACTTCTGCGCAAGGAATGATTCGATG 660
601 CAAGCTCTGGCATCATCCACTGCTGGCTCAGCACTTCTGCGCAAGGAATGATTCGATG 660
661 TGGGCTATTCCATCAAAACGCAATGATTCGACGCTCGGAGCTTCTTCTTTGATTCACAA 720
661 TGGGCTATTCCATCAAAACGCAATGATTCGACGCTCGGAGCTTCTTCTTTGATTCACAA 720
721 ATCGCTGCTCGGTGCGAATCAGCTCAGTATAGCTTTTCCGCGCGCGCTGCTTCGCG 780
721 ATCGCTGCTCGGTGCGAATCAGCTCAGTATAGCTTTTCCGCGCGCGCTGCTTCGCG 780
781 TCGTCTTACGTCGCGCTGTTTCAACAGCTTCCACGATGCTAGACCACTCTTTTACAA 840
781 TCGTCTTACGTCGCGCTGTTTCAACAGCTTCCACGATGCTAGACCACTCTTTTACAA 840
841 GTTGTTCGAAGCAGCGGCTTTGTTATCCGTTTCAGACGTTAGCGGTTGAAATTTATCAAGA 900
841 GTTGTTCGAAGCAGCGGCTTTGTTATCCGTTTCAGACGTTAGCGGTTGAAATTTATCAAGA 900
901 GCGAGCTCGGTTATGCTCCGAGCTCAGCTCAGCGGTAAACCGGTAGTTCATCGTCA 960
901 GCGAGCTCGGTTATGCTCCGAGCTCAGCTCAGCGGTAAACCGGTAGTTCATCGTCA 960
961 ATTGCAACAACAACGACGACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAACA 1020
961 ATTGCAACAACAACGACGACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAACA 1020
1021 GAGCTTCACGAGTTCATGAGACACCAACAGCAGCGGTATCGAACCTGA 1071
1021 GAGCTTCACGAGTTCATGAGACACCAACAGCAGCGGTATCGAACCTGA 1071

RESULT 2

US-09-938-842A-1034
; Sequence 1034, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1034
LENGTH: 1071
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1034

Query Match 100.0%; Score 1071; DB 12; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCAATTCAGAGCTTGAAGAGCTTGCAGGCAAGATCAAACTCTAAGACCGGTT 60
DB 1 ATGGCGCAATTCAGAGCTTGAAGAGCTTGCAGGCAAGATCAAACTCTAAGACCGGTT 60
QY 61 GATCTAACCATCATCAACGGCTCAGAAAGCTGCAAACTTCAAGACCTTTCCAGTAAT 120
DB 61 GATCTAACCATCATCAACGGCTCAGAAAGCTGCAAACTTCAAGACCTTTCCAGTAAT 120
QY 121 CCCACAGTGAAGTCTCGAGCCCAAGCGGAGCGGTGATGCGGTCTTTCAATGTCCTTA 180
DB 121 CCCACAGTGAAGTCTCGAGCCCAAGCGGAGCGGTGATGCGGTCTTTCAATGTCCTTA 180
QY 181 GCTCCACCGCTTTCGACAGGACCACTTGAAGAGAGCTTCGACTTAAAGACCGTCAACG 240
DB 181 GCTCCACCGCTTTCGACAGGACCACTTGAAGAGAGCTTCGACTTAAAGACCGTCAACG 240
QY 241 AAGGTTGAAGGAGAGGAGAGATGAGTACGATGCTGCGACGCTGCGGTGCGGTGCGATTTT 300
DB 241 AAGGTTGAAGGAGAGGAGAGATGAGTACGATGCTGCGACGCTGCGGTGCGGTGCGATTTT 300
QY 301 CAATTAATCTCGAGAGTGTAGTCTCAAAATCCGACGCGCAACGATTCGCTGCTGCTGAG 360
DB 301 CAATTAATCTCGAGAGTGTAGTCTCAAAATCCGACGCGCAACGATTCGCTGCTGCTGAG 360
QY 361 AACGCTGAGCCGCGATTTATAGCCGCAACGGGTACGGAACGCTTCCCGCATCGCCATG 420
DB 361 AACGCTGAGCCGCGATTTATAGCCGCAACGGGTACGGAACGCTTCCCGCATCGCCATG 420
QY 421 TCGGTTAAGCAACCTTAAATAATCCGACGACGACGATTCGATTCGATGCGGTGAA 480
DB 421 TCGGTTAAGCAACCTTAAATAATCCGACGACGACGATTCGATTCGATGCGGTGAA 480
QY 481 AATCTGATGAGAGAGAGACGTAACGCTTCTAACAGTGAATATAGACATAGCGAC 540
DB 481 AATCTGATGAGAGAGAGACGTAACGCTTCTAACAGTGAATATAGACATAGCGAC 540
QY 541 GCCGTTTCAGCTTCTCGGTTTATAGTCTCAATTCGACGACGACGATCCCAACCTCCG 600
DB 541 GCCGTTTCAGCTTCTCGGTTTATAGTCTCAATTCGACGACGACGATCCCAACCTCCG 600
QY 601 CAAGCTCTGGCATCATCCACTGCTGGCTCAGCACTTCTGCGCAAGGAATGATTCGATG 660
DB 601 CAAGCTCTGGCATCATCCACTGCTGGCTCAGCACTTCTGCGCAAGGAATGATTCGATG 660
QY 661 TGGGCTATTCCATCAAAACGCAATGATTCGACGCTCGGAGCTTCTTCTTTGATTCACAA 720
DB 661 TGGGCTATTCCATCAAAACGCAATGATTCGACGCTCGGAGCTTCTTCTTTGATTCACAA 720
QY 721 ATCGCTGCTCGGTGCGAATCAGCTCAGTATAGCTTTTCCGCGCGCGCTGCTTCGCG 780

Db 721 ATGCTGGTCCGTCGAATCAGCCTCAGTTATTAGCTTTTCCGCGCGCGCTGCTTCGCG 780
Y 781 TCGTCTTACGTCGCCGCTGTTCACACAGGCTTCCAGATGGCTAGACCACTCCTTTTACA 840
b 781 TCGTCTTACGTCGCCGCTGTTCACACAGGCTTCCAGATGGCTAGACCACTCCTTTACA 840
Y 841 GTTGTTCCAGACAGCGGCTTGTATCCGTTTCCAGATGGCTAGACCACTCCTTTACA 900
b 841 GTTGTTCCAGACAGCGGCTTGTATCCGTTTCCAGATGGCTAGACCACTCCTTTACA 900
Y 901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGGCTTAACACCGGTAGTTTCATCGTCA 960
b 901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGGCTTAACACCGGTAGTTTCATCGTCA 960
Y 961 ATTGCAACACAGCAGCAGCTGCTAGAGACTTCTCCCTAGAGATATACGAGAAACA 1020
b 961 ATTGCAACACAGCAGCAGCTGCTAGAGACTTCTCCCTAGAGATATACGAGAAACA 1020
Y 1021 GAGCTTCCAGCTTATGAGCAGCAGCTCAAGCTCAGGCTTAACACCGGTAGTTTCATCGTCA 1071
b 1021 GAGCTTCCAGCTTATGAGCAGCAGCTCAAGCTCAGGCTTAACACCGGTAGTTTCATCGTCA 1071

RESULT 3

IS-09-924-035A-502/c
Sequence 502, Application US/09924035A
Patent No. US20020142319A1
GENERAL INFORMATION:
APPLICANT: Glach, Jm
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2011US
CURRENT APPLICATION NUMBER: US/09/924,035A
CURRENT FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: US 60/148,784
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 900
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 502
LENGTH: 460
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(460)
OTHER INFORMATION: n = A,T,C or G
US-09-924-035A-502

Query Match 41.8%; Score 448; DB 10; Length 460;
Best Local Similarity 99.6%; Pred. No. 8e-145;
Matches 459; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Y 600 GCAAGCTCGGCATCATCAGTCAAGTATCCAGCGTCCGAGCTTCTTCTTGAATCCACA 659
b 460 GCAAGCTCGGCATCATCAGTCAAGTATCCAGCGTCCGAGCTTCTTCTTGAATCCACA 401
Y 660 GTGGGCTATTCATCAAGCAAGTATCCAGCGTCCGAGCTTCTTCTTGAATCCACA 719
b 400 GTGGGCTATTCATCAAGCAAGTATCCAGCGTCCGAGCTTCTTCTTGAATCCACA 341
Y 720 AATCGTGTGCTCGATCAGCTCAGTATTAAGCTTTCCGCGCGCGCTGCTCGCC 779
b 340 AATCGTGTGCTCGATCAGCTCAGTATTAAGCTTTCCGCGCGCGCTGCTCGCC 282
Y 780 GTGCTCTTACGTCGCGCTGTTCACAGCTTCCAGATGGCTAGACCACTCCTTTTACA 839
b 281 GTGCTCTTACGTCGCGCTGTTCACAGCTTCCAGATGGCTAGACCACTCCTTTTACA 222
Y 840 AGTGTGTCCAGACAGCGCTTGTATCCGTTTTCAGACCTTACGAGCTTTCAGATTTATCAAG 899
b 221 AGTGTGTCCAGACAGCGCTTGTATCCGTTTTCAGACCTTACGAGCTTTCAGATTTATCAAG 162

QY 900 AGGACGTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGGTAGTTTCATCGTC 959
Db 161 AGGACGTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGGTAGTTTCATCGTC 102
QY 960 AATTGCAACACAGCAGCAGCTGCTAGAGACTTCTCCCTAGAGATATACGAGAAACA 1019
Db 101 AATTGCAACACAGCAGCAGCTGCTAGAGACTTCTCCCTAGAGATATACGAGAAACA 42
QY 1020 AGAGCTTCCAGCTTATGAGCAGCAGCTCAAGCTCAGCGGTAT 1060
Db 41 AGAGCTTCCAGCTTATGAGCAGCAGCTCAAGCTCAGCGGTAT 1

RESULT 4

US-09-770-444-615/c
Sequence 615, Application US/09770444
Patent No. US2002023280A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jm
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krieger, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2027 (PABA-016PRV)
CURRENT APPLICATION NUMBER: US/09/770,444
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,502
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 615
LENGTH: 453
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(453)
OTHER INFORMATION: n = A,T,C or G
US-09-770-444-615

Query Match 41.0%; Score 439; DB 9; Length 453;
Best Local Similarity 99.1%; Pred. No. 1.1e-141;
Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 607 CTGGCATCATCAGTCTGCTCAGCAACTTCTGCGCGAAGAAATGATCCGATGTGGGT 666
Db 453 CTGGCATCATCAGTCTGCTCAGCAACTTCTGCGCGAAGAAATGATCCGATGTGGGT 394
QY 667 ATTCCATCAACGCAATGATTCGAGCGTTCGAGCTTCTTCTTGAATCCACAAATCGCT 726
Db 393 ATTCCATCAACGCAATGATTCGAGCGTTCGAGCTTCTTCTTGAATCCACAAATCGCT 334
QY 727 GTTCCGTCAATCAGCTCAGTTATTAAGCTTTCCGCGCGCTGCTTCGCGCTGCT 786
Db 333 GTTCCGTCAATCAGCTCAGTTATTAAGCTTTCCGCGCGCTGCTTCGCGCTGCT 275
QY 787 TAGTTCGCGCTGTTTCAACAGCTTCCAGATGGCTAGACCACTTCTTACAAATGTT 846

Db 274 TACGTCGCGCTTTTCAACAGCTTCCAGATGGCTAGACACCTCTTTTCAAGTTGTT 215
QY CCAAGCAGCGGTTTGTATCCGTTTTCAGACGTTTTCGAATTTATCAAGACGACG 906
Db 214 CCAAGCAGCGGTTTGTATCCGTTTTCAGACGTTTTCGAATTTATCAAGACGACG 155
QY TCGGTTATGGCTCCGAGCTCAGCTCAGCGGTACCAACCGGTAGTCTATCGTCAATTGCA 966
Db 154 TCGGTTATGGCTCCGAGCTCAGCTCAGCGGTACCAACCGGTAGTCTATCGTCAATTGCA 95
QY 967 ACAACACGACGACACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAAACAAGAGCTT 1026
Db 94 ACAACACGACGACACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAAACAAGAGCTT 35
QY 1027 CACAGTTCATGAGACACACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAAACAAGAGCTT 1060
Db 34 CACAGTTCATGAGACACACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAAACAAGAGCTT 1

RESULT 5

US-09-770-696-257
; Sequence 257, Application US/09770696
; Patent No. US2001004940A1

GENERAL INFORMATION:

; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthews, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: Thaliana
; FILE REFERENCE: 2031US (PARA-020PRV)
; CURRENT APPLICATION NUMBER: US/09/770,696
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/178,278
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 911
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257
; LENGTH: 185
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana

US-09-770-696-257

Query Match 17.3%; Score 185; DB 9; Length 185;

Best Local Similarity 100.0%; Pred. No. 1.8e-53;

Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 GCAAGATCAAACTTAAAGACGCTTGAATCAACATCATCAACGCGGTGAGAAAGTCG 94
Db 1 GCAAGATCAAACTTAAAGACGCTTGAATCAACATCATCAACGCGGTGAGAAAGTCG 60
QY 95 AAACCTCAAGACCTTTTCAAGTAATCCACAGTAGTCTCGAGCCCAAGGCGGCGG 154
Db 61 AAACCTCAAGACCTTTTCAAGTAATCCACAGTAGTCTCGAGCCCAAGGCGGCGG 120
QY 155 TGATGCCGCTGTTTTCATGTCTTTAGTCTCAGCTTCCAGACGCTTTCAGACGACCATTTGAAGA 214

Db 121 TGATGCCGCTGTTTTCATGTCTTTAGTCTCAGACGCTTTCAGACGACCATTTGAAGA 180
QY 215 GAGCT 219
Db 181 GAGCT 185

RESULT 6

US-09-934-455-169
; Sequence 169, Application US/09934455
; Publication No. US20030121070A1

GENERAL INFORMATION:

; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omlira

TITLE OF INVENTION: Genes for Modifying Plant Traits IV

; FILE REFERENCE: MBI-0025

; CURRENT APPLICATION NUMBER: US/09/934,455

; PRIOR FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/227439

; PRIOR FILING DATE: 2000-08-22

; PRIOR APPLICATION NUMBER: MBI-0022

; PRIOR FILING DATE: 2001-11-16

; PRIOR APPLICATION NUMBER: MBI-0023

; PRIOR FILING DATE: 2001-04-17

; NUMBER OF SEQ ID NOS: 516

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 169

; LENGTH: 1728

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (106)...(1575)

; OTHER INFORMATION: G1064

; US-09-934-455-169

Query Match 12.0%; Score 129; DB 11; Length 1728;

Best Local Similarity 71.0%; Pred. No. 2.4e-33;

Matches 172; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 189 GTCTTCGACGACCAACATTTGAAGAGAGCTTCCACTAAAGACCGTCAACGAGGTTGA 248
Db 420 GGCAGCTAAAAGGCCACCGTTGAAACGAGCGTCCGAAAGACGACACGAAAGTAGA 479
QY 249 AGGAAGAGGAGAGATACGATGCGCTGCCACGCTGCGCGTAGGATTTTCAATTAAC 308
Db 480 CGAAGAGGAGAGAGATACGATGCGCGCTGCGCGTAGGATTTTCACTTCTTAA 539
QY 309 TCAGAGTAGTTCACAAATCCGACGCGGAAACGATTCCGCTGTTGAGAGACGCTGA 368
Db 540 GCGAGAGTAGTTCATAAATCCGACGCGTGAACAATAGAGTGGCTTCTTCAACAAGCTGA 599
QY 369 GCGGCGGATTATAGCCGACGCGGTACGGGAACGTTTCCGCCATCCCGATGCGGTAA 428
Db 600 ACCATCTGTAAATCGCCGCCACCGGAACCGGAACCAATCCCGCGGAATTTCACTTCTTAA 659
QY 429 C 429
Db 660 C 660

RESULT 7

US-10-225-068-155

Sequence 165, Application US/10225068
Publication No. US20030217383A1
GENERAL INFORMATION:
APPLICANT: Mendel Biotechnology, Inc.
APPLICANT: Reuber, T. Lynne
APPLICANT: Riechmann, Jose Luis
APPLICANT: Heard, Jacqueline E.
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Adam, Luc J.
APPLICANT: Dubell, Arnold T.
APPLICANT: Ratcliffe, Oliver
APPLICANT: Pineda, Omaira
APPLICANT: Yu, Guo-Liang
APPLICANT: Broun, Pierre E.
TITLE OF INVENTION: STRESS-RELATED POLYNUCLEOTIDES AND
POLYPEPTIDES IN PLANTS
FILE REFERENCE: 51442002040
CURRENT APPLICATION NUMBER: US/10/225,068
CURRENT FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: 60/310,847
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 60/336,049
PRIOR FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/338,692
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: 10/171,468
PRIOR FILING DATE: 2002-06-14
NUMBER OF SEQ ID NOS: 246
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 165
LENGTH: 1728
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (106)... (1575)
US-10-225-068-165

Query Match 12.0%; Score 129; DB 13; Length 1728;
Best Local Similarity 71.0%; Pred. No. 2.4e-33;
Matches 171; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
Y 189 GTCTTCGACGAGCACCACCTTGAAGAGCTTCGACTAAGACCGCTCACGAAAGTTGA 248
D 420 GCGAGTAAAGCCACCCCTTGAAGAGCTTCGACTAAGACCGCTCACGAAAGTTGA 479
Y 249 AGAAGAGGAGGAGATACGATCGCTCCACGCTGCGGCTAGATTTTCAATTAAAC 308
D 480 CGAAGAGGAGGAGATACGATCGCTCCACGCTGCGGCTAGATTTTCAATTAAAC 539
Y 309 TCAGAGTTAGGTCAAAATCCGAGCGGCGAAACGATTGCGTGTGTTGGAGAACGCTGA 368
D 540 CGAGAGTAGGTCAAAATCCGAGCGGCGAAACGATTGCGTGTGTTGGAGAACGCTGA 599
Y 369 GCGGCGATTATAGCGCGACCGGCTACGGAACGGTTCCGCCATCGGCATGTCGGTTAA 428
D 600 ACCATCTGTAATCGCGCGACCGGCTACGGAACGGTTCCGCCATCGGCATGTCGGTTAA 659
Y 429 C 429
D 560 C 560

RESULT 8
US-09-934-455-137
Sequence 137, Application US/09934455
Publication No. US20030121070A1
GENERAL INFORMATION:
APPLICANT: Adam, Luc
APPLICANT: Creelman, Robert
APPLICANT: Dubell, Arnold
APPLICANT: Heard, Jacqueline
APPLICANT: Jiang, Cai-Zhong

APPLICANT: Keddle, James
APPLICANT: Pilgrim, Marsha
APPLICANT: Ratcliffe, Oliver
APPLICANT: Reuber, Lynne
APPLICANT: Riechmann, Jose Luis
APPLICANT: Yu, Guo-Liang
APPLICANT: Pineda, Omaira
TITLE OF INVENTION: Genes for Modifying Plant Traits IV
FILE REFERENCE: MBI-0025
CURRENT APPLICATION NUMBER: US/09/934,455
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/227439
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: MBI-0022
PRIOR FILING DATE: 2001-11-16
PRIOR APPLICATION NUMBER: MBI-0023
PRIOR FILING DATE: 2001-04-17
NUMBER OF SEQ ID NOS: 516
SOFTWARE: PatentIn version 3.1
SEQ ID NO 137
LENGTH: 1231
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (107)... (1084)
OTHER INFORMATION: G805
US-09-934-455-137

Query Match 11.6%; Score 124.6; DB 11; Length 1231;
Best Local Similarity 65.0%; Pred. No. 6.5e-32;
Matches 184; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
Y 202 CCACCATTTGAAGAGCTTCGACTAAGACCGCTCACGAAAGTTGAAGAGGAGGAGA 261
D 239 CCTCCTCTAAACGAACTCTACTAAGACCGCTCACGAAAGTTGAAGAGGAGGAGC 298
Y 262 AGGATACGGATGCTCCACGCTGCGGCTAGATTTTCAATTAACTCGAGAGTTAGGT 321
D 299 CGATCGGTATGCTGCCATGTGTGTCACGTGCTTTTCACTCACAGTGGCTTGT 358
Y 322 CACAAATCCGACGGCGAAACGATTCGGTGTGTTGGAGAACGCTGAGCGCGGATTATA 381
D 359 CACAAATCCGACGGCGTAAACTATTGAGTGGCTACTCCACAAAGCAGAACCCAGCG 418
Y 382 GCGCGCACGGGTACGGGAAACGGTTCCGCCATCGCTGCTTAAACGGAACCTTAA 441
D 419 GCGCTACAGGAGTGGAAACCTTCGGCTAACTCACTTTTAAACATCTCCTTCT 478
Y 442 ATCCGACGACGAGAACGCTGATTCGATGATGGGTGAAATC 484
D 479 AGCTCAAGATCTTCTCTCTGCTGCTCATCTTCTGTAACAATC 521

RESULT 9
US-10-302-267-61
Sequence 61, Application US/10302267
Publication No. US20030229915A1
GENERAL INFORMATION:
APPLICANT: Keddle, James
APPLICANT: Fromm, Michael
APPLICANT: Heard, Jacqueline
APPLICANT: Riechmann, Jose Luis
APPLICANT: Adam, Luc
APPLICANT: Broun, Pierre
APPLICANT: Pineda, Omaira
APPLICANT: Reuber, Lynne
APPLICANT: Zhang, James
APPLICANT: Yu, Guo-Liang
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Samaha, Raymond
APPLICANT: Pilgrim, Marsha
APPLICANT: Creelman, Robert

; TITLE OF INVENTION: PLANT GENE SEQUENCES II
; FILE REFERENCE: MBI-0007
; CURRENT APPLICATION NUMBER: US/10/302,267
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US/09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,880
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: 60/121,037
; PRIOR FILING DATE: 1999-02-22
; PRIOR APPLICATION NUMBER: 60/124,278
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/129,450
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/135,134
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 60/144,153
; PRIOR FILING DATE: 1999-07-15
; PRIOR APPLICATION NUMBER: 60/161,143
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/162,656
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 61
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(1081)
; OTHER INFORMATION: G805
US-10-302-267-61

Query Match 11.6%; Score 124.6; DB 12; Length 1231;
Best Local Similarity 65.0%; Pred. No. 6.5e-32;
Matches 184; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 202 CCACCATTTGAAGAGAGGCTTCGACTAAAGACCGTCACACGAGGTTGAAGGAGGAGGAGA 261
DB 239 CCTCCTCTTAACGACCTCTACTTAAGACCGACACACCAACGTCGAGGCCGCGGT 298
QY 262 AGGATACGGATGCTGCGACGTCGTGCGGTAGGATTTTCAATTAATCTCAGAGTTAGGT 321
DB 299 CGGATCCGCTATGCTGCGCATGTGCTGCGAGTGTCTTTCAGCTCACAACGTGAGCTTGGT 358
QY 322 CACAAATCCGACGCGGAAACGATTCCGTTGTTGGAGAACGTCGACGCGGATTATA 381
DB 359 CACAAATCCGACGCGTGAACCTATTGAGTGGCTACTCCACAGAGAACGAGCGGTTATA 418
QY 382 GCGGCCACGGGTACGGGAACGGTTCCCGCCATGCCATGTCGGTTAACGGAACTTAAAA 441
DB 419 GCGGCTACAGGACTGGAACCAATTCGGCTAACTTCTCTTAAACATCTCAGTTGGT 478
QY 442 ATCCGACGACGACGACGCTGATTCTGATATGGTGAAATC 484
DB 479 AGCTCAAGATCTTCTCTCTGCTGCTCATCTTCGTACAACTC 521

RESULT 10
US-10-225-068-245
; Sequence 245, Application US/10225068
; Publication No. US20030217383A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Adam, Luc J.
; APPLICANT: Dubell, Arnold T.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pineda, Omaira

; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E.
; TITLE OF INVENTION: STRESS-RELATED POLYNUCLEOTIDES AND
; TITLE OF INVENTION: POLYPEPTIDES IN PLANTS
; FILE REFERENCE: 51442002040
; CURRENT APPLICATION NUMBER: US/10/225,068
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 246
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 245
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(1084)
US-10-225-068-245
Query Match 11.6%; Score 124.6; DB 13; Length 1231;
Best Local Similarity 65.0%; Pred. No. 6.5e-32;
Matches 184; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 202 CCACCATTTGAAGAGAGGCTTCGACTAAAGACCGTCACACGAGGTTGAAGGAGGAGGAGA 261
DB 239 CCTCCTCTCTTAACGACCTCTACTTAAGACCGACACACCAACGTCGAGGCCGCGGT 298
QY 262 AGGATACGGATGCTGCGACGTCGTGCGGTAGGATTTTCAATTAATCTCAGAGTTAGGT 321
DB 299 CGGATCCGCTATGCTGCGCATGTGCTGCGAGTGTCTTTCAGCTCACAACGTGAGCTTGGT 358
QY 322 CACAAATCCGACGCGGAAACGATTCCGTTGTTGGAGAACGTCGACGCGGATTATA 381
DB 359 CACAAATCCGACGCGTGAACCTATTGAGTGGCTACTCCACAGAGAACGAGCGGTTATA 418
QY 382 GCGGCCACGGGTACGGGAACGGTTCCCGCCATGCCATGTCGGTTAACGGAACTTAAAA 441
DB 419 GCGGCTACAGGACTGGAACCAATTCGGCTAACTTCTCTTAAACATCTCAGTTGGT 478
QY 442 ATCCGACGACGACGACGCTGATTCTGATATGGTGAAATC 484
DB 479 AGCTCAAGATCTTCTCTCTGCTGCTCATCTTCGTACAACTC 521

RESULT 11
US-10-295-403-147
; Sequence 147, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519

;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: 60/101,349
;; PRIOR FILING DATE: 1998-09-22
;; PRIOR APPLICATION NUMBER: 60/103,312
;; PRIOR FILING DATE: 1998-10-06
;; PRIOR APPLICATION NUMBER: 60/108,734
;; PRIOR FILING DATE: 1998-11-17
;; PRIOR APPLICATION NUMBER: 60/113,409
;; PRIOR FILING DATE: 1998-12-22
;; NUMBER OF SEQ ID NOS: 170
;; SOFTWARE: Patent In Ver. 2.0
;; SEQ ID NO 147
;; LENGTH: 1604
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (143)..(1345)
;; OTHER INFORMATION: G802
JS-10-295-403-147

Query Match 11.4%; Score 121.6; DB 15; Length 1604;
Best Local Similarity 64.6%; Pred. No. 8.5e-31;
Matches 181; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
2Y 188 CGCTTCGACAGACACCACTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAGGTTG 247
2b 285 CATCTACGGCAACGGCGACTACACGAAGGGTCGACTAAAGACCGTCACACGAAAGTCG 344
2Y 248 AAGAAAGAGGAGAGGATACCGATCCCTGCCACGCTGTCGGCTAGGATTTTCAATTAA 307
2b 345 ACGAAAGAGACGCTCGGATTCGTATCGCGCGCTATGTGCAGCTAGAGTTTTTCAGTTAA 404
2Y 308 CTCGAGAGTGTGTACAACTCGACGGCGAAACGATTCGGTGTGTTGGAGACGCTG 367
2b 405 CGAGAGAGTGTGTACAACTCGATCGGCGCGCTATGTGCAGCTAGAGTTTTTCAGTTAA 464
2Y 368 AGCCGGCGATTTATAGCGCGACACGGGTACGGGAACGGTTCGCCCATCGCCATGTCGGTTA 427
2b 465 AGCCAGCTATTGTTGCTGTCTACAGGTACAGGAATTCGGCGGAATCTCTCTACTTTAA 524
2Y 428 ACGAACTTTAAATCCCGACGACGACGACGCTGATTC 467
2b 525 GTGTTTCGTTACGAAGTAGTGGATCGACTCTCTCTGCTCC 564

RESULT 12

US-09-770-961-675
;; Sequence 675, Application US/09770961
;; Publication No. US20030115639A1
;; GENERAL INFORMATION:
;; APPLICANT: Gorlach, Jörn
;; APPLICANT: An, Yong-Qiang
;; APPLICANT: Hamilton, Carol M.
;; APPLICANT: Price, Jennifer L.
;; APPLICANT: Raines, Tracy M.
;; APPLICANT: Yu, Yang
;; APPLICANT: Rameaka, Joshua G.
;; APPLICANT: Page, Amy
;; APPLICANT: Matthew, Abraham V.
;; APPLICANT: Ledford, Brooke L.
;; APPLICANT: Woessner, Jeffrey P.
;; APPLICANT: Haas, William David
;; APPLICANT: Garcia, Carlos A.
;; APPLICANT: Kricker, Maja
;; APPLICANT: Slader, Ted
;; APPLICANT: Davis, Keith R.
;; APPLICANT: Allen, Keith
;; APPLICANT: Hoffman, Neil
;; APPLICANT: Hurban, Patrick
;; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
;; TITLE OF INVENTION: thaliana
;; FILE REFERENCE: 2026 (PARA-015PRV)

;; CURRENT APPLICATION NUMBER: US/09/770,961
;; CURRENT FILING DATE: 2001-01-26
;; PRIOR APPLICATION NUMBER: 60/178,466
;; PRIOR FILING DATE: 2000-01-27
;; NUMBER OF SEQ ID NOS: 999
;; SOFTWARE: Pab:SEQ for Windows Version 4.0
;; SEQ ID NO 675
;; LENGTH: 490
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)..(490)
;; OTHER INFORMATION: n = A,T,C or G
US-09-770-961-675

Query Match 11.0%; Score 118; DB 11; Length 490;
Best Local Similarity 73.3%; Pred. No. 7.2e-30;
Matches 151; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 211 AAGAGAGCTTCGACTAAAGACCGTCACACGAGGTTGAAGGAGAGGAGGATACGG 270
Db 249 AAGAGAGAGCTCTAACAAAGACAGACACACTAAAGTCGAAAGGTAGAGGTCCGGAATTCGG 308
QY 271 ATGCTCTCCACGCTGTCGGCTAGGATTTTCAATTAACTCGAGAGTTAGTCACAATCC 330
Db 309 ATGCTCTCTTTTGTCTGCTAGGATTTTCAATTGCTAGAGATTTGGTCAATAATCT 368
QY 331 GACGGCGAAACGATTCGGTGTGTTGGAGAACGCTGAGCCGGCGANTTAGCCGCCAG 390
Db 369 GATGCTGAACACTATCCAGTGGCTGCTTCAACAGCTGAGCCATCGATTTATTCAGCTACT 428
QY 391 GGTACGGGACGGTTCGCCCATCGC 416
Db 429 GGTTCAGAACTATACGGCTCTGTC 454

RESULT 13

US-09-934-455-133
;; Sequence 133, Application US/09934455
;; Publication No. US20030121070A1
;; GENERAL INFORMATION:
;; APPLICANT: Adam, Luc
;; APPLICANT: Creelman, Robert
;; APPLICANT: Dubell, Arnold
;; APPLICANT: Heard, Jacqueline
;; APPLICANT: Jiang, Cai-Zhong
;; APPLICANT: Keddle, James
;; APPLICANT: Pilgrim, Marsha
;; APPLICANT: Ratcliffe, Oliver
;; APPLICANT: Reuber, Lynne
;; APPLICANT: Riechmann, Jose Luis
;; APPLICANT: Yu, Guo-Liang
;; APPLICANT: Pineda, Omaira
;; TITLE OF INVENTION: Genes for Modifying Plant Traits IV
;; FILE REFERENCE: MBI-0025
;; CURRENT APPLICATION NUMBER: US/09/934,455
;; CURRENT FILING DATE: 2001-08-22
;; PRIOR APPLICATION NUMBER: 60/227439
;; PRIOR FILING DATE: 2000-08-22
;; PRIOR APPLICATION NUMBER: MBI-0022
;; PRIOR FILING DATE: 2001-11-16
;; PRIOR APPLICATION NUMBER: MBI-0023
;; PRIOR FILING DATE: 2001-04-17
;; NUMBER OF SEQ ID NOS: 516
;; SOFTWARE: Patent In version 3.1
;; SEQ ID NO 133
;; LENGTH: 1009
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (27)..(746)

OTHER INFORMATION: G801
US-09-934-455-133

Query Match 10.4%; Score 111.2; DB 11; Length 1009;
Best Local Similarity 71.6%; Pred. No. 2.7e-27;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 209 TGAAGAGAGCTTCGACTAAACCGCTCACACGAGGTTGAGGAGGAGGAGGAGGATAC 268
DB 103 TTAAGAAACCCCGCTTAAGATCGACACAAAGTTGACGAGAGGAGGAGGATTC 162
QY 269 GGATGCTGCCACGTGTGCGGTAGGATTTTCAATTAACCTCGAGAGTTAGGTACAAAT 328
DB 163 GTATGCCAATCATTTTGGCAGCTCGAGTTTTTCAATTGACAGAGAGTTAGGTACAAAGT 222
QY 329 CCGACGGCGAAACGATTTCGGTGGTTTGGAGAACGCTGAGCCGCGCATATAGCCGCCA 388
DB 223 CCGATGGTCAACACCATAGAGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCA 282
QY 389 CGGGTACGGGAACGGTTCGCCGCCA 412
DB 283 CTGGAACCTGGCACTACTCCGGCGA 306

RESULT 14

US-10-225-068-135
Sequence 135, Application US/10225068
Publication No. US20030217383A1
GENERAL INFORMATION:
APPLICANT: Mendel Biotechnology, Inc.
APPLICANT: Reuber, T. Lynne
APPLICANT: Riechmann, Jose Luis
APPLICANT: Heard, Jacqueline E.
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Adam, Luc J.
APPLICANT: Dubell, Arnold T.
APPLICANT: Ratcliffe, Oliver
APPLICANT: Pineda, Omalra
APPLICANT: Yu, Guo-Liang
APPLICANT: Brown, Pierre E.
TITLE OF INVENTION: STRESS-RELATED POLYNUCLEOTIDES AND
FILE REFERENCE: 514442002040
CURRENT APPLICATION NUMBER: US/10/225,068
CURRENT FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: 60/310,847
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 60/336,049
PRIOR FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/338,692
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: 10/171,468
PRIOR FILING DATE: 2002-06-14
NUMBER OF SEQ ID NOS: 246
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 135
LENGTH: 1009
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (27)...(746)
US-10-225-068-135

Query Match 10.4%; Score 111.2; DB 13; Length 1009;
Best Local Similarity 71.6%; Pred. No. 2.7e-27;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 209 TGAAGAGAGCTTCGACTAAACCGCTCACACGAGGTTGAGGAGGAGGAGGATAC 268
DB 103 TTAAGAAACCCCGCTTAAGATCGACACAAAGTTGACGAGAGGAGGAGGATTC 162
QY 269 GGATGCTGCCACGTGTGCGGTAGGATTTTCAATTAACCTCGAGAGTTAGGTACAAAT 328

DB 163 GTATGCCAATCATTTTGGCAGCTCGAGTTTTTCAATTGACAGAGAGTTAGGTACAAAGT 222
QY 329 CCGACGGCGAAACGATTTCGGTGGTTTGGAGAACGCTGAGCCGCGCATATAGCCGCCA 388
DB 223 CCGATGGTCAACACCATAGAGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCA 282
QY 389 CGGGTACGGGAACGGTTCGCCGCCA 412
DB 283 CTGGAACCTGGCACTACTCCGGCGA 306
RESULT 15
US-10-295-403-145
Sequence 145, Application US/10295403
Publication No. US20030101481A1
GENERAL INFORMATION:
APPLICANT: Heard, Jacqueline
APPLICANT: Riechmann, Jose Luis
APPLICANT: Adam, Luc
APPLICANT: Brown, Pierre
APPLICANT: Pineda, Omalra
APPLICANT: Reuber, Lynne
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Keddle, James
APPLICANT: Zhang, James
APPLICANT: Benito, Maria-Ines
APPLICANT: Yu, Guo-Liang
APPLICANT: Fromm, Mike
TITLE OF INVENTION: PLANT GENE SEQUENCES I
FILE REFERENCE: MBI-0003
CURRENT APPLICATION NUMBER: US/10/295,403
CURRENT FILING DATE: 2002-11-15
PRIOR APPLICATION NUMBER: US/09/394,519
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: 60/101,349
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/103,312
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/108,734
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/113,409
PRIOR FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 170
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 145
LENGTH: 1009
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (27)...(743)
OTHER INFORMATION: G801
US-10-295-403-145

Query Match 10.4%; Score 111.2; DB 15; Length 1009;
Best Local Similarity 71.6%; Pred. No. 2.7e-27;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 209 TGAAGAGAGCTTCGACTAAACCGCTCACACGAGGTTGAGGAGGAGGAGGATAC 268
DB 103 TTAAGAAACCCCGCTTAAGATCGACACAAAGTTGACGAGAGGAGGAGGATTC 162
QY 269 GGATGCTGCCACGTGTGCGGTAGGATTTTCAATTAACCTCGAGAGTTAGGTACAAAT 328
DB 163 GTATGCCAATCATTTGCGCAGCTCGAGTTTTTCAATTGACAGAGAGTTAGGTACAAAGT 222
QY 329 CCGACGGCGAAACGATTTCGGTGGTTTGGAGAACGCTGAGCCGCGCATATAGCCGCCA 388
DB 223 CCGATGGTCAACACCATAGAGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCA 282
QY 389 CGGGTACGGGAACGGTTCGCCGCCA 412

Db 283 CTGGAACTGGCACTACTCCGGCGA 306

Search completed: February 2, 2004, 01:10:23
Job time : 414.599 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:43:46 ; Search time 2477.95 Seconds
(without alignments)
10504.671 Million cell updates/sec

Title: US-09-938-842A-1034

Perfect score: 1071
Sequence: 1 atgggacaaatcagaagct.....caggtcatcgaaaccactga 1071

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	409.6	38.2	798	28	BH511920 BOHGA14TF
2	382	35.7	486	9	AV825375 AV825375
3	360.8	33.7	373	10	BES22768 M28F8STM
4	314	29.3	335	14	T88453 12149 Lambd

BH450567	BQGD041TR	556	28	BH450567
B35068	ATT33707	388	14	B35068
BE523004	M31B65TM	378	10	BE523004
BH543331	BOGNC02TF	725	28	BH543331
BH470366	BOHC133TF	680	28	BH470366
CB292965	UCRCS01_0	791	14	CB292965
AU239448	AU239448_0	634	9	AU239448
AW697442	NF118G06S	652	9	AW697442
CA929907	QTCU2CA.P4	439	14	CA929907
BQ865644	QO52SH15.Y	747	13	BQ865644
BF070046	st63s02.Y	649	10	BF070046
AI973663	sd08a06.Y	459	9	AI973663
BM437189	VVA015FL1	716	12	BM437189
BQ874162	QF4J066.Y	673	13	BQ874162
BE440597	SP47e10.Y	626	10	BE440597
BF278295	GA_EB003	661	10	BF278295
BH741132	gt27g10_9	586	28	BH741132
AV442079	AV442079	514	9	AV442079
BM522413	sak98f11.	566	12	BM522413
BQ403313	GA_EB005	668	13	BQ403313
BZ460479	BOMXW58TR	736	29	BZ460479
BH866140	BONJAW58TR	773	28	BH866140
AW649873	EST328327	671	9	AW649873
AJ503095	AJ503095	522	9	AJ503095
AW695867	NF100A09S	649	9	AW695867
AW695783	NF098G07S	657	9	AW695783
BQ442740	GA_EB001	907	10	BQ442740
BQ405430	GA_EB008	693	13	BQ405430
BF269051	GA_EB000	652	10	BF269051
BZ519530	BOMQX23TR	728	29	BZ519530
CB972449	CAB30001	775	14	CB972449
BH664697	BOHUX48TR	785	28	BH664697
BF052551	EST437781	416	10	BF052551
BG447197	GA_EB004	830	10	BG447197
AJ489125	AJ489125	422	9	AJ489125
BI932752	EST526641	527	12	BI932752
BE463339	EST354562	562	10	BE463339
BG446467	GA_EB003	581	10	BG446467
BG445804	GA_EB002	815	10	BG445804
BM111907	EST559443	519	12	BM111907
BE344238	EST5409400	622	10	BE344238

ALIGNMENTS

RESULT 1
BH511920/c
LOCUS BH511920 798 bp DNA linear GSS 13-DEC-2001
DEFINITION BOHGA14TF BOHG Brassica oleracea genomic clone BOHGA14, genomic survey sequence.
ACCESSION BH511920
VERSION BH511920.1 GI:17720010
KEYWORDS GSS
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustoids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 798)
AUTHORS Town.C.D., Van Aken,S., Utterback.T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other GSSs: BOHGA14TR
Contact: Chris Town.
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: ctown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

```
FEATURES
source
Location/Qualifiers
1..798
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOHGA14"
/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 161 a 210 c 225 g 202 t
ORIGIN
Query Match 38.2%; Score 409.6; DB 28; Length 798;
Best Local Similarity 80.3%; Pred. No. 1.9e-118;
Matches 525; Conservative 0; Mismatches 84; Indels 45; Gaps 2;
QY 184 CCACGCTCTCCACAGGACCACTTGAAGAGAGCTTCGACTTAAGACCGTCACACGAG 243
Db 798 CCGTCTGCTCAATAGGACCGCACTGAAGAGAGCTTCGACTTAAGACCGTCACACGAG 739
QY 244 GTTGAAGGAGGAGGAGAGGATACGGATGCTGCCAGCTGTGCGGTAGGATTTTCAA 303
Db 738 GTCAAGGAGGAGGAGGAGGATACGGATGCTGCCAGCTGTGCGGTAGGATTTTCCAG 679
QY 304 TTAACCTCGAGAGTTAGTTCACAAATCCGACGCGCAACGATTCGGTGTGTTGGAGAC 363
Db 678 TTGACTCGAGAGCTTGGTCACAAATCCGACGCGGAGAGATTCGGTGTGTTGGAGAAC 619
QY 364 GGTGAGCGCGCGGATATAGCCGCCAGGTACGGGAACGGTTCGCCGCATCGCATGTGCG 423
Db 618 GCGGAGCGCGCGGATATCGCCGCCAGGTACGGGAACGGTTCGCCGCATCGCATGTGCG 559
QY 424 GTTAACGGAACCTTAAATATCCGACGAGAGAGAGAGCTGATTCGATGATGGTGAATA 483
Db 558 GTCAACGGAACCTTAAATATCCGACGAGAGAGAGAGCTGATTCGATGATGGTGAATA 499
QY 484 CTGATGAAGAGAAAGTAAAGACCTTCTACAGTGAATATAGATATAGATAGGAGCGCC 543
Db 498 CCGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 439
QY 544 GTTTCAGCTTCCTCCGTTTATGCTTCAATTTGCCAGAGAGAGAGAGAGAGAGAGAGAG 603
Db 438 GTCTCCGTTTCTCCGTTTATGCTCCG----- 411
QY 604 GCTCTGGCATCATCACTTGTGCTGAGCAACTTCTGCGCAGAGAGATGATCCGATGTGG 663
Db 410 -----CATCCACCCGCGCTCAACAAACGCTGCGCGCAGAGAGATGATCCGATGTGG 361
QY 664 GCTATTCATCAAAAGCAATGATTCGACGCTGCGAGAGTTCCTTCTGATTCACAAATC 723
Db 360 GGTATTCATCAAAAGCAATGATTCGACGCTGCGAGAGTTCCTTCTGATTCACAAAGCA 301
QY 724 GCTGTGCTCGAATCAGCTCAGTTATAGTTTTC---CGCGCGCGCTGCTTCCGCG 780
Db 300 GTGCGCGCTGCAACAGAGCTCAGATATAGTTTATCCGCGCGCGCTGCTGCTTACCG 241
QY 781 TCGTCTTACGCTGCGCGCTTCAACAGGCTTCACAGTGGCTAGACCACTTCCT 834
Db 240 TCGACGTAATGCTGCGCGCTTCAGAGGCTTCTTATGCTAGACCACTTCCT 187
RESULT 2
AV825375 486 bp mRNA linear EST 01-APR-2002
LOCUS AV825375 RAFL7 Arabidopsis thaliana cDNA clone RAFL07-08-P04 5',
DEFINITION mRNA sequence.
ACCESSION AV825375
VERSION AV825375.1 GI:19867435
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsi.s.
1 (bases 1 to 486)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@tc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998) cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda phage-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.
Location/Qualifiers
1..486
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL07-08-P04"
/dev_stage="rosette plants"
/lab_host="DH10B"
/clone_lib="RAFL7"
/note="Site 1: BamHI; Site 2: SalI; subjected to
cold-treated (1, 2, 5, 10, 24 hr)"
BASE COUNT 131 a 118 c 117 g 117 t
ORIGIN
Query Match 35.7%; Score 382; DB 9; Length 486;
Best Local Similarity 99.2%; Pred. No. 8.4e-110;
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGACAAATTCAGAAGCTTGAAGAAGTTGAGCAAGATCAAACTCTAAGAGCCGTT 60
Db 102 ATGGGACAAATTCAGAAGCTTGAAGAAGTTGAGCAAGATCAAACTCTAAGAGCCGTT 161
QY 61 GATCTAACCATCATCAACGGCTCAGAACGTCGAACTTCAGACCTTCCAGTAAT 120
Db 162 GATCTAACCATCATCAACGGCTCAGAACGTCGAACTTCAGACCTTCCAGTAAT 221
QY 121 CCCACAGTAGTCTCGAGCCCAAGCGGAGCGCGTGATGCCGTCGTTTCAATGCTTTA 180
Db 222 CCCACAGTAGTCTCGAGCCCAAGCGGAGCGCGTGATGCCGTCGTTTCAATGCTTTA 281
QY 181 GCTCCACCGTCTTCGACAGACCACTTGAAGAGAGCTTCGACTTAAGACCGTCACAG 240
Db 282 GCTCCACCGTCTTCGACAGACCACTTGAAGAGAGCTTCGACTTAAGACCGTCACAG 341
QY 241 AAGGTTGAAGGAGAGGAGAGAGGATACGATGCTCCGTCGCTGCGGTAGGATTTT 300
Db 342 AAGGTTGAAGGAGAGGAGAGAGGATACGATGCTCCGTCGCTGCGGTAGGATTTT 401
QY 301 CAATTAACTCGAGAGTTAGTTCACAAATCCGACGCGGAAACGATTCGGTGTGTTGAG 360
Db 402 CAATTAACTCGAGAGTTAGTTCACAAATCCGACGCGGAAACGATTCGGTGTGTTGAG 461
QY 361 AAGCTGTAGCGCGGATTTATAGCCG 385
Db 462 AAGCTGTAGCGCGGATTTATAGCCG 486
RESULT 3
BE522768 373 bp mRNA linear EST 19-MAR-2001
LOCUS BE522768
```

DEFINITION M28F8STM Arabidopsis developing seed Arabidopsis thaliana cDNA
 Clone M28F8 5', mRNA sequence.
 BE522768
 BE522768.1 GI:9780746
 EST.
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 (bases 1 to 373)
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
 Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
 A new set of Arabidopsis expressed sequence tags from developing
 seeds. The metabolic pathway from carbohydrates to seed oil
 Plant Physiol. 124 (4), 1582-1594 (2000)
 20567808
 11115876
 PUBMED
 COMMENT
 Contact: Benning, C
 Dept. of Biochemistry & Molecular Biology
 Michigan State University
 224 Biochemistry, Michigan State University, East Lansing, MI 48824
 USA
 Tel: 517 355 1609
 Fax: 517 353 9334
 Email: benning@msu.edu
 Michigan State University DNA Sequencing Facility Arabidopsis
 Biological Resource Center, The Ohio State University, 309 Botany &
 Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
 6142920603 TEL: 6142929371.
 Location/Qualifiers
 1. .373
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="M28F8"
 /tissue_type="seed"
 /dev_stage="5-13 days after flowering"
 /lab_host="E.coli"
 /note="Organ: Developing seed"
 /clone_lib="Arabidopsis developing seed"
 /note="Organ: Developing seed; Vector: pBluescript SK-"
 Site_1: EcoRI; Site_2: XhoI
 73 a 118 c 82 g 100 t

BASE COUNT
 ORIGIN
 Query Match 33.7%; Score 360.8; DB 10; Length 373;
 Best Local Similarity 99.5%; Pred. No. 4e-103; 2; Indels 0; Gaps 0;
 Matches 362; Conservative 0; Mismatches 0; Gaps 0;
 534 AAGCGACGCGGTTTACGCTTCCTCCGGTTTAGCTCCAAATGGCCAGCAGCAACGATCCA 593
 10 AGCGACGCGGTTTACGCTTCCTCCGGTTTAGCTCCAAATGGCCAGCAGCAACGATCCA 69
 594 ACCTCCGACGCTCGGCATCATCCACTGGCTGGCTAGCACTTCTGCCAGGAGATGTA 653
 70 ACCTCCGACGCTCGGCATCATCCACTGGCTGGCTAGCACTTCTGCCAGGAGATGTA 129
 654 TCCGATGTGGGCTATTCCATCAACCAATGATTCGACGGTGGGAGCTTCTTCTTGAT 713
 130 TCCGATGTGGGCTATTCCATCAACCAATGATTCGACGGTGGGAGCTTCTTCTTGAT 189
 714 TCCACAAATCGCTGGTCCGTCGGAATCAGCTCAGTTATTAGCTTTCCGCGCGCGCTGC 773
 190 TCCACAAATCGCTGGTCCGTCGGAATCAGCTCAGTTATTAGCTTTCCGCGCGCGCTGC 249
 774 TTCCGCTCGCTTACGTTCGCGCTGTTCAACAGGCTTCCAGATGGCTAGACCACTCC 833
 250 TTGCGCGCTGCTTACGTTCGCGCTGTTCAACAGGCTTCCAGATGGCTAGACCACTCC 309
 834 TTTCAGATGTTTCCAGCAGCGGCTTTGATCCGTTTCAGACGTTAGCGGTTGCAATT 893
 310 TTTTACAGATGTTTCCAGCAGCGGCTTTGATCCGTTTCAGACGTTAGCGGTTGCAATT 369

QY 894 ATCA 897
 DB 370 ATCA 373

RESULT 4
 T88453
 LOCUS 12149 Lambda-PRI2 Arabidopsis thaliana cDNA clone 156J17, mRNA
 DEFINITION
 T88453
 T88453.1 GI:936289
 EST
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 (bases 1 to 335)
 Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh
 , L., Ohlrogge, J., Raikhe, N., Somerville, S., Thomashow, M., Retzel
 , E. and Somerville, C.
 Genes Galore: a summary of methods for accessing results from
 large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 Plant Physiol. 106, 1241-1255 (1994)
 95148729
 PUBMED
 7846151
 COMMENT
 On Apr 14, 1993 this sequence version replaced gi:716805.
 Contact: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
 Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 22313tcm@bm.cl.msu.edu
 Seq primer: T7 dye primer.
 Location/Qualifiers
 1. 335
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="var Columbia"
 /db_xref="taxon:3702"
 /clone="156J17"
 /clone_lib="Lambda-PRI2"
 /note="Vector: Lambda Zip-Lox; Site 1: Sal; Site 2: Not;
 Lambda PRI2 is a cDNA library derived from equal
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7
 day germinated etiolated seedlings; 2) tissue culture
 grown roots; 3) staged plants half with 24 hour light
 cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
 same plants as 3 but aerial tissue (stems, flowers and
 siliques). The vector is BRL's lambda Zip-Lox. The cDNA
 inserts were directionally cloned with Sal-Not arms using
 oligo dT primed cDNA."
 96 a 75 c 82 g 69 t 13 others

BASE COUNT
 ORIGIN
 Query Match 29.3%; Score 314; DB 14; Length 335;
 Best Local Similarity 94.6%; Pred. No. 2.8e-88;
 Matches 317; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 QY 3 GGCACAAATTCAGAGCTTGAAAGTTGCAGCAAGATCAAACTCTAAGAGCCGTGA 62
 DB 1 GGCACAAATTCAGAGCTTGAAAGTTGCAGCAAGATCAAACTCTAAGAGCCGTGA 60
 QY 63 TCTAACCATCATCAACGGCTCAGAACCTCGAACTTCAAGACCTTCCAGTAATCC 122
 DB 61 TCTAACCATCATCAACGGCTCAGAACCTCGAACTTCCAGACCTTCCAGTAATCC 120
 QY 123 CACAGTGAAGTCTCGAGCCCAAGCGGAGCGGGTGAAGTCCCGTGGTTTCAATGTCCTTAC 182


```

Db      121  CACAGTGTCTCGAGCCCAAGCGGACGTNNATGCGTCNTTTTCAATGTTTATG 180
QY      183  TCCACCGCTTCGACAGACCAACATTGAAGAGAGCTTCGACTAAAGACCGTCAACGAA 242
Db      181  TCCACCGTNTTCGACAGGACCAACATTGAAGAGAGCTTCGACTAAAGACCGTCAACGAA 240
QY      243  GGTTCAGAGAGAGGAGGAGGATGATCGGATCGCTGCGACGCTGTCGGCTAGGATTTTCA 302
Db      241  GGTTCAGAGAGAGGAGGAGGATGATCGGATCGCTGCGACGCTGTCGGCTAGGATTTTCA 300
QY      303  ATTAAGTCTAGAGTGTAGGTACAAATCCGACGGCG 337
Db      301  ATTAAGTCTAGAGTGTAGGTACAAATCCGACGGCG 335

RESULT 5
BH450567/c
LOCUS      BH450567 656 bp DNA linear GSS 12-DEC-2001
DEFINITION BOGDQ41TR BOGD Brassica oleracea genomic clone BOGDQ41, genomic survey sequence.
ACCESSION  BH450567
VERSION    BH450567.1 GI:17636278
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 656)
            Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
            Whole genome shotgun sequencing of Brassica oleracea
            Unpublished
            Other GSSs: BOGDQ41TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: Sheared ends.
            Location/Qualifiers
                1..656
                /organism="Brassica oleracea"
                /mol_type="genomic DNA"
                /strain="T0100DH3"
                /db_xref="taxon:3712"
                /clone="BOGDQ41"
                /clone_lib="BOGD"
                /note="Vector: pHD1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHD1 using BstXI linkers"
BASE COUNT 142 a 167 c 190 g 157 t
ORIGIN
Query Match 29.1%; Score 311.2; DB 28; Length 656;
Best Local Similarity 78.4%; Pred. No. 3.2e-87;
Matches 431; Conservative 0; Mismatches 73; Indels 46; Gaps 3;

QY      288  GGCTAGGATTTTCAATTAATCGAGAGTTAGGTACAAATCCGACGGCGGAAACGATTTCG 347
Db      656  GGCTAGGATTTTCAATTAATCGAGAGTTAGGTACAAATCCGACGGCGGAGAGATTTCG 597
QY      348  GTGGTTGTGGAGAACGTCAGCCGCGGATATAGCCGCCACGGGTACGGGAACGGTTCC 407
Db      596  GTGGTTG-TGGAGAACGGCGAGCGCGCGGATAATCGCCGCCACGGGTACGGGAACGGTTCC 538
QY      408  CCCCATCGCATGTCGGTTAAAGGAACCTTAAATATCCGACGACGACGACGTCGATTC 467
Db      537  CCCCATCGCATGTCGGTCACGGAACGCTTAAATATCCGACGACGACGACGTCGATTC 478
QY      468  TGATATGGGTGAAATCTGATGAAGAGAAACGTTAAACGACCTTCTTACAGTGTAGTAT 527

```

```

Db      477  TGAGTGGCGGAAAAATCCGGGAAGAGACGTAAGGACCTTCTACCACTGTAGTATAT 418
QY      528  AGACATAAGCGAGCGGCTTTCAGCTTCCTCGGTTTACCTCAATTCGCCAGCAAC 587
Db      417  AGACATAAGCAAGACGGTCTCGTTTCTCGGTTTACCTCC----- 376
QY      588  GATCCAACTCCGCAAGCTCTGGCATCATCCACTGTGGCTCAGCAACTTCTGCGCAAGG 647
Db      375  -----GGCATCCACCGCGCTCAACAACGCTTCGCCCAGG 340
QY      648  AATGATATCCGATGTGGGTATTTCATCAACGCAATGATTCCGACGGTCGAGGTTTCTT 707
Db      339  GATGATTCGATGTGGGCTTTCATCAACGCAATGATTCCGACGGTCGAGGTTTCTT 280
QY      708  CTTGATTCACAATCGCTGGTCCGTAATCAGCTCAGTATATAGCTTTC---CCGC 764
Db      279  CTTGTTCAAAACGTAAGTGGCCCGTCGAACGAGCTCAGATATTAGTTTATCCGCGCGC 220
QY      765  CGCGCTGCTTCGCGCTGCTTTACGTCGCGCTGTTTCAACAGGCTTCCAGATGGCTAG 824
Db      219  CGCTGCTGCTTTACCGTCGACGATGTGCGCGCTTTCAGCAGGCTTCTTCTATGGCTAG 160
QY      825  ACCACTCTCT 834
Db      159  ACCACTCTCT 150

RESULT 6
Z35068
LOCUS      Z35068 388 bp mRNA linear EST 18-JUL-1994
DEFINITION ATTS3707 Gif-SeedA Arabidopsis thaliana cDNA clone YAI344 5', mRNA sequence.
ACCESSION  Z35068
VERSION    Z35068.1 GI:510816
KEYWORDS   EST.
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 388)
            CNRS
            The Arabidopsis thaliana transcribed genome: the GDR cDNA program
            Unpublished
            Contact: Berthomieu P., Guerrier D., Giraudat J.
            Genetique Moleculaire d'Arabidopsis
            ISV - UPR40, CNRS
            Avenue de la Terrasse, 1198 Gif-sur-Yvette Cedex, France
            Email: Giraudat@cnrs-gif.fr.
            Location/Qualifiers
                1..388
                /organism="Arabidopsis thaliana"
                /mol_type="mRNA"
                /db_xref="taxon:3702"
                /clone="YAV344"
                /clone_lib="Gif-SeedA"
                /note="Vector: Lambda ZAPII non-oriented; Physiological condition: greenhouse plants. tissue type: Green silique"
BASE COUNT 102 a 94 c 108 g 83 t 1 others
ORIGIN
Query Match 27.3%; Score 292.6; DB 14; Length 388;
Best Local Similarity 97.8%; Pred. No. 1.9e-81;
Matches 317; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY      86  GAAAGCTCGAAATCTCAAGACCTTTCGAAGTAAATCCACAGTGAAGTCTCGAGCCCAAGG 145
Db      1  GAAAGCTCGAAATCTCAAGACCTTTCGAAGTAAATCCACAGTGAAGTCTCGAGCCCAAGG 60
QY      146  CGAGCGCGTGTATGCGCTGCTTTCAATGCTTTAGCTCCAGCGCTTTCGACGAGGCCAC 205
Db      61  CGAG-CGTGAATGCGCTGCTTTCAATGCTTTAGCTCCAGCGCTTTCGACGAGGCCAC 119

```

206 CATTGAAGAGAGCTTCGACTAAGACCGTCAACGAGAGTTGAAGGAAGAGGAGGA 265
 120 CATTGAAGAGAGCTTCGACTAAGACCGTCAACGAGAGTTGAAGGAAGAGGAGGA 179
 266 TAGGATGCTCCACGCTGCGGTAGGATTTTCAATTAACGAGAGTTAGTCA 325
 180 TAGGATGCTCCACGCTGCGGTAGGATTTTCAATTAACGAGAGTTAGTCA 239
 326 AATCCGAGCGCAACGATTCGGTGTGTGAGAACGCTGAGCGCGGATTA-TAGCC 384
 240 AATCCGAGCGCAACGATTCGGTGTGTGAGAACGCTGAGCGCGGATTA-TAGCC 299
 385 GCACGGGTACGGGACGCTTCCC 408
 300 GCACGGGTACGGGACGCTTCCC 323

RESULT 7
 E523004
 LOCUS
 DEFINITION
 BE523004 Arabidopsis developing seed Arabidopsis thaliana cDNA
 clone M31B6 5', mRNA sequence.
 BE523004
 BE523004.1 GI:9780982
 EST.
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 1 (bases 1 to 378)
 White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
 Ilarduya, O., Jaworski, J.G., Ohlrogge, J. and Benning, C.
 A new set of Arabidopsis expressed sequence tags from developing
 seeds. The metabolic pathway from carbohydrates to seed oil
 Plant Physiol. 124 (4), 1582-1594 (2000)
 20567808
 11115876
 Contact: Benning, C
 Dept. of Biochemistry & Molecular Biology
 Michigan State University
 224 Biochemistry, Michigan State University, East Lansing, MI 48824
 , USA
 Tel: 517 355 1609
 Fax: 517 353 9334
 Email: benning@msu.edu
 Michigan State University DNA Sequencing Facility Arabidopsis
 Biological Resource Center, The Ohio State University, 309 Botany &
 Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
 6142920603 TEL: 6142929371.

FEATURES
 source
 1. 378
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="M31B6"
 /tissue_type="seed"
 /dev_stage="5-13 days after flowering"
 /lab_host="E.coli"
 /clone_lib="Arabidopsis developing seed"
 /note="Organ: Developing seed; Vector: pBluescript SK-;
 Site 1: EcoRI; Site 2: XhoI"
 108 a 97 c 81 g 92 t
 Query Match 23.8%; Score 255.2; DB 10; Length 378;
 Best Local Similarity 95.3%; Pred. No. 1.5e-69;
 Matches 263; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 1 ATGGCGACAAATTCAGAGCTTGAAGAGTTGACGAGCAAGATCAAACTCTAAGACCGCTT 60

103 ATGGCGACAAATTCAGAGCTTGAAGAGTTGACGAGCAAGATCAAACTCTAAGACCGCTT 162
 61 GATCTTAACCATCATCAACGCGCTCAGAAACGTCGAACTTCAAGACCTTTCAAGTAAT 120
 163 GATCTTAACCATCATCAACGCGCTCAGAAACGTCGAACTTCAAGACCTTTCAAGTAAT 222
 121 CCACAGTGTGCTCGAGCCCAAGCGGAGCGGTGATGCGCTTTTCAATGCTTTTA 180
 223 CCACAGTGTGCTCGAGCCCAAGCGGAGCGGTGATGCGCTTTTCAATGCTTTTA 282
 181 GCTCCACCGCTTTCGACAGGACCAACATTTGAAGAGAGCTTTCGACTAAAGACCGTCACACG 240
 283 GCTCCACCGCTTTCGACAGGACCAACATTTGAAGAGAGCTTTCGACTAAAGACCGTCACACG 342
 241 AAGTTGAAGGAAGGAGGAGGATACGATGCGCT 276
 343 AAGGTTGAAGGAAGGATGAAGGATACGATGCGCT 378

RESULT 8
 BH543331
 LOCUS
 DEFINITION
 BH543331 BOGC Brassica oleracea genomic clone BOGCN02, genomic
 survey sequence.
 BH543331
 BH543331.1 GI:17795112
 GSS.
 Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 1 (bases 1 to 725)
 Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished
 Other_GSSs: BOGCN02TR
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-8208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 1. 725
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOGCN02"
 /clone_lib="BOGC"
 /note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into PHOS1 using BstXI linkers"
 182 a 198 c 191 g 154 t
 BASE COUNT
 ORIGIN
 Query Match 18.2%; Score 195.2; DB 28; Length 725;
 Best Local Similarity 73.9%; Pred. No. 2.5e-50;
 Matches 309; Conservative 0; Mismatches 63; Indels 46; Gaps 3;
 420 GTCCGTTAAGGAGCACTTAAATCCGAGCAGCAGCAGCGCTGATTTGATATGGTGA 479
 26 GTCCGTTAAGGAGCACTTAAATCCGAGCAGCAGCGCTGATTTGATATGGTGA 84
 480 AATCTGATGAGGAAGCAAGCAAGTAAACGACCTTCAACAGTGAATATAGACATAGCGA 539
 85 AATCTGAGGAAGCAAGCAAGTAAACGACCTTCAACAGTGAATATAGACATAGCGA 144
 540 CGCCGTTTTCAGCTTCTCCGGTTTAGCTTCAATTCGACGACCAACGATCCACCTCC 599

```

Db 145 ACGGTCTCCGTTTCTCCGCTTATAGTC----- 173
QY 600 GCAAGCTCTGGCATCATCCACTGTGCGCTCAGCAACTCTCGCGCAAGAAATGTATCCGAT 659
Db 174 -----CGGATTCACCGCGCTCACAAACGCTGCGCGCAGGATATCCGAT 222
QY 660 GTGGCTATTCATCAAAACGCAATGATTCCGACGGTGGAGCTTCTCTTGTATCCACA 719
Db 223 GTGGCGGTTTCCATCAAAACGCGAGTGTTCGACGCTTGGAGCTTCTCTTGTGTCAAAA 282
QY 720 AATCCCTGTGTCGTGATCAGCTCAGCTTATAGCTTTTC---CCGCGCGCGTGTTC 776
Db 283 CGTACTGCGCGCTCGAACAGCTCAGATATAGCTATCCGCGCGCGTGTGCTTT 342
QY 777 GCCGTGCTTACGTGCGCGCTGTGTTCACAGGCTTCCACGATGCTAGACCACTCT 834
Db 343 ACCGTGAGCTATGTGCGCGCGCTTCCAGAGCTTCTCTATGCTAGACCACTCT 400

RESULT 9
BH470366/c
LOCUS BH470366 680 bp DNA linear GSS 13-DEC-2001
DEFINITION BOHC133TF BOHC Brassica oleracea genomic clone BOHC133, genomic survey sequence.
ACCESSION BH470366
VERSION BH470366.1 GI:17678477
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 680)
AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other GSSs: BOHC133TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: Sheared ends.

FEATURES             Location/Qualifiers
     source            1..680
     /organism="Brassica oleracea"
     /mol_type="genomic DNA"
     /strain="TO1000DH3"
     /db_xref="taxon:3712"
     /clone="BOHC133"
     /clone_lib="BOHC"
     /note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT          145 a 188 c 175 g 172 t

Query Match          15.4%; Score 165; DB 28; Length 680;
Best Local Similarity 72.5%; Pred.No. 9.5e-41;
Matches 263; Conservative 0; Mismatches 55; Indels 45; Gaps 2;

QY 475 GGTGAATCTGATGAAGAAGAAAGTAACGACCTTCTAACAGTGAAGTATATAGACATA 534
Db 668 GCGGAAATCCGCGGAAGAGAGAGAGTAAGGACCTTCTACAGTGAATATAGACATA 609
QY 535 ACGGACGCGTTTCAGCTTCTCCCGGTTAGCTTCCAAATGCGACGACCAACGATCAA 594
Db 608 AACGAACCGGCTCCGTTCTCCCGGTTAGCTCCG----- 572
QY 595 CTTCCGCAAGCTCTGGCATCATCCACTGTGGCTCAGCAACTTCTGCCGCAAGGATGAT 654

```

```

Db 571 -----CATCACCGCGCTCAACAAACGCTCCGCGCAGGATGATT 531
QY 655 CCGATGTGGCTATTCCATCAAAACGATGATTCCGACGGTCGAGCTTCTTCTTGATT 714
Db 530 CCGATGTGGCGTTTCCATCAAAACGAGTGTTCGACGGTTGGAGCTTCTTCTTGTT 471
QY 715 CCACAAATCGCTGTGCTCGTGAATCAGCTCAGTTATTAGCTTTTC---CGCGCGCGCT 771
Db 470 CAAACGCTAGCTGGCGCTCGAAACGAGCTCAGATATTAGCTTATCCGCGCGCGCTGCT 411
QY 772 GTTTCGCGCTGCTTCTAGCTGCGCTGTTCACAGGCTTCCAGATGCTAGACCACT 831
Db 410 GTTTACGCTGACGATGTCGCCCTTCAGCAGGCTTCTTATGCTAGACCACT 351

832 CCT 834
350 CCT 348

RESULT 10
CB292965
LOCUS CB292965 791 bp mRNA linear EST 28-FEB-2003
DEFINITION UCRC01_05bd03_g1 Washington Navel orange cold acclimated flavedo & albedo cDNA library Citrus sinensis cDNA clone UCRC01_05bd03, mRNA sequence.
ACCESSION CB292965
VERSION CB292965.1 GI:28618422
KEYWORDS EST.
SOURCE Citrus sinensis
ORGANISM Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE 1 (bases 1 to 791)
AUTHORS Close, R.J., Collura, X., Fenton, R.D., Feuerbacher, O., Kim, H.R., Kudrna, D., Wanmaker, S., Wing, R. and Yu, Y.
TITLE Development of EST Resources and New Genetic Markers for California Citrus
JOURNAL Unpublished
COMMENT Contact: Timothy Close
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124
Tel: 9097873318
Fax: 9097874437
Email: timothy.close@ucr.edu
Seq primer: T3.

FEATURES             Location/Qualifiers
     source            1..791
     /organism="Citrus sinensis"
     /mol_type="mRNA"
     /cultivar="Washington Navel"
     /db_xref="taxon:2711"
     /clone="UCRC01_05bd03"
     /tissue_type="Rind containing flavedo and albedo"
     /dev_stage="Mature fruit"
     /lab_host="E. coli TUC121"
     /clone_lib="Washington Navel orange cold acclimated flavedo & albedo cDNA library"
     /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the field at University of California, Riverside Agricultural Operations since 1983. The scion was Washington Navel orange and the rootstock Carizzo Citrange. Tissue from mature fruit was harvested at mid-day in January 2002 during a cold spell, when pre-dawn temperatures were approximately -2 to 2 degrees C. Approximately 2 cm median sections of the rind were excised in the field from several fruits, then wrapped in aluminum foil and frozen quickly in dry ice. Total RNA was extracted using a phenol extraction procedure described in J. Japanese Soc. Hort. Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA library was made, and 1 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript

```

SK(-) phagemids. All steps to this point were performed in the TU Close lab at the University of California, Riverside (Fenton). Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI350 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wanmaker) using the Harvest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

```

BASE COUNT      203 a 230 c 192 g 166 t
ORIGIN
Query Match      15.4%; Score 164.4; DB 14; Length 791;
Best Local Similarity 71.2%; Pred. No. 1.6e-40;
Matches 232; Conservative 0; Mismatches 91; Indels 3; Gaps 1;

2y 214 AGAGCTTCGACTAAGACCGTCAACGAGGTTGAGGAGGAGGAGGAGGATGATG 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2b 56 AGAGCGCTTCCAAAGATCGGCACCAAGTGCAGAGCGCGCGCGTAGGATTAGA 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2y 274 CCTGCCACGCTGTCGGGTAGGATTTTCAATTAACCTCGAGAGTTAGGTCACAAATCCGAC 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2b 116 CCGCCACGCTGTCGGCCAGGATCTTTCACCTAACCCGCGAATCGGTCAAGTCCGAC 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2y 334 GCGGAACAGTTCGGTGGTGTGGAGAACGCTGAGCGCGCGGATATAGCCGCCAGGGT 393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2b 176 GGTGAACACCATCCGTTGGTTACTTGAACACCGCCAGCCAGCCATTTATTGAGGCCACTGCG 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2y 394 ACGGGAACGGTTCGCCCATCGCATGTGGTTAAACGGAACCTTAAATATCCGACGAGC 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2b 236 ACTGGCACGCTGCCCGTATCGCGTTCGGTTAACGGAACGTTAAAGATTCCCAACG 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2y 454 ACGAACGCTGATCTGATATGGTGAATCTGATGAAGAAACGTTAAACGACCTTCT 513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2b 296 TCTCCAGCGAAACCTG---CCGCGAAGAACTGCCGAGGAACCGCGAAGAGACCGTGT 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2y 514 AACAGTGAGTATATACATACGGA 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2b 353 AACAGCGAGTTCATTGATTTAACGA 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
LOCUS      AU239448
DEFINITION AU239448 RAFL19 Arabidopsis thaliana cDNA clone RAFL19-81-E01 5',
            mRNA sequence.
ACCESSION  AU239448
VERSION     AU239448
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 634)
AUTHORS     Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,
            Itoh,M., Ichi,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu
            ,M., Hayashizaki,Y. and Shinozaki,K.
            Large scale analysis of Arabidopsis full-length cDNA
            Unpublished
            Contact: Motoaki Seki
            RIKEN Genomic Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4339
            Fax: 81-298-36-9060
            Email: mseki@rtt.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially

```

as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified plasmid vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES

Location/Qualifiers
 1..634
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL19-81-E01"
 /tissue_type="mixture of silique and flower"
 /lab_host="pE10B"
 /clone_lib="RAFL19"
 /notes="Site 1: BamHI; Site 2: SalI; Subtraction Library"
 BASE COUNT 187 a 137 c 174 g 133 t 3 others
 ORIGIN

```

Query Match      14.7%; Score 157.8; DB 9; Length 634;
Best Local Similarity 75.9%; Pred. No. 1.8e-38; Indels 0; Gaps 0;
Matches 195; Conservative 0; Mismatches 62;

QY 195 GACAGGACCAACCTTGAAGAGAGCTTCGACTAAAGACCGTCAACGAGGTTGAAGAAAG 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 GACTAAACCGGCTCCGAGAGAGACCGACTTCTAAAGACCGTCAACGAAAGTAGAGGACG 246
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 AGGAGAGAGATAGGATGCTCCAGGTGCGGTAGGATTTTCAATTAACTCGAGA 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 AGTCCGAGGATCCGAATCGCGCGGTTCGCTGCTCGGGTCTTTCAATTGACCGGTA 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 315 GTTAGGTCACAAAATCCGACCGCGGAAACGATTCCGTGTTGTTGGAAACCGCTGAGCGCGC 374
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 ACTTGTCACAAATCCGACCGGAGAACGATACGTTGTTATTGGACGAGTGAACCGGC 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 375 GATTATAGCGCCACGCGGTACGGGAACGGTTCCCGCCATCGCCATCGTGGTTAAACGAAAC 434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 GATAATTGAAGCAACCGGAACCGGAACCTGTACCGGCTATTGCTGTATCGTTTAAACGGAAC 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 435 CTTAAAAATCCCGACGA 451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 TTTAAAAATCCCGACGA 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 12
LOCUS      AW697442
DEFINITION AW697442 652 bp mRNA linear EST 15-JUN-2000
            NF11806ST1E1051 Developing stem Medicago truncatula cDNA clone
            NF11806ST 5', mRNA sequence.
ACCESSION  AW697442
VERSION     AW697442
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
            Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
            Medicago.
            1 (bases 1 to 652)
            He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
            ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
            ,R.A.
            Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
            Unpublished
            Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert length: 652 Std Error: 0.00

```


Plate: 118 row: G column: 06
Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers
1. .652

FEATURES
source
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NFI1806ST"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/clone_lib="Developing stem"
/note="Vector: Lambda Zap; Contains a mixture of
intermodal stem segments."
BASE COUNT 221 a 190 C 119 G 122 t
ORIGIN

Query Match 14.4%; Score 154.6; DB 9; Length 652;
Best Local Similarity 72.9%; Pred. No. 1.9e-37;
Matches 199; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 180 AGCTCCACCGTCTTCGACAGGACCCACCATTTGAAGAGAGCTTCGACTAAGACCGTCCACAC 239
Db 228 AGCTCAGCTCAGCCCAACAAACATTTCAAAGAGATCATCTACAAAGACCGTCCACAC 287
QY 240 GAAGCTTGAAGAGAGGAGAGATACGATGCTCCGACGTCGCTGCGGCTAGGATTTT 299
Db 288 AAGAGTCAGAGGTCGAGCGCGAGGATCCGATGCTCCGACATGTCAGCTCGGATCTT 347
QY 300 TCAATTAATCGAGAGTTAGTCACAAATCCGAGCGGAGAAACGATTCGGTGGTTGGGA 359
Db 348 TCAGCTACCGGAGAACTCGGACATAGTCCGATGCGGAAACATACGCTGGCTCCGCA 407
QY 360 GAAGCTGTAGCGCGCGGATTTAGCGCGCCAGGTCAGGAAACGTTCCCGCCATGCCAT 419
Db 408 ACAGCGGAGCCAGCTATCATTCGCGCCACCGGTACCGGACATCCGACCATGCCAT 467
QY 420 GTGCTTAAACGGAACCTTAAATCCCGACGAC 452
Db 468 GTCAGTCAACGGAACACTAAATATCCCCACCAC 500

RESULT 13
CA929907/c
LOCUS
DEFINITION MTU2CA.P4.C10 Aspen apex cDNA Library Populus tremuloides cDNA,
mRNA sequence.
ACCESSION CA929907.1 GI:27418388
VERSION CA929907
KEYWORDS EST
SOURCE Populus tremuloides (quaking aspen)
ORGANISM Populus tremuloides
REFERENCE 1 (bases 1 to 439)
AUTHORS Ranjan, P., Kao, Y.-Y., Harding, S.A., Jiang, H., Joshi, C.P. and Tsai, C.-J.
TITLE Expressed sequence tags from Aspen
JOURNAL Unpublished
COMMENT Contact: Tsai C-J
Plant Biotech Research Center
Michigan Technological University, School of Forest Resources &
Environmental Science
1400 Townsend Drive, Houghton, MI 49931-1295, USA
Tel: 906 487 2914
Fax: 906 487 2915
Email: chtsai@mtu.edu.
Location/Qualifiers
1. .439

FEATURES
source
/organism="Populus tremuloides"
/mol_type="mRNA"
/db_xref="taxon:3693"
/clone_lib="Aspen apex cDNA Library"

BASE COUNT 91 a 98 C 126 G 124 t
ORIGIN

Query Match 13.9%; Score 149.2; DB 14; Length 439;
Best Local Similarity 62.0%; Pred. No. 7.9e-36;
Matches 254; Conservative 0; Mismatches 153; Indels 3; Gaps 1;
QY 145 CGGAGCGCGTGATGCGCGTCTTTTCAATGTCTTTAGCTCCACCGTCTTCGACAGGACCA 204
Db 435 GCTATGCAGATGCAGACGCCAATGCCAATTTCAATGCTCTGCCAGTCACTACACAGCA 376
QY 205 CCATTGAAGAGAGCTTCGACTAAGACCGTTCACAGAGGTTGAAGAGAGGAGGAGG 254
Db 375 ACAAACGAGCGCGATCCACCAAGACCGCCACACAAAGTGGAGGCGCGTGGCCGTAGG 316
QY 265 ATAGGATGCGCTGCGCAGCTGTGCGGCTAGGATTTTCAATTAACCTCGAGAGTTAGGTAC 324
Db 315 ATCCGATACCGCTACTTGTGCTGCGCGGATCTTCCAATTGACCGGAGATTAGGCCAT 256
QY 325 AATCCGACGCGGAAAGATTCGGTGTGTTGAGAGAACCTGAGCGCGGCGGATATAGCC 384
Db 255 AATCAGACGCGTGAACCGGTAGATGCTCTTGAACATCTGAGGAAGCTATTATTGAA 196
QY 385 GCCACGGGTACGGGAACGCTTCCGCGCATGCCATGCGGTAAACGGAACCTTAAAAATC 444
Db 195 GCAACCGGACCGGACGCTCCCTGCTATCGCGTCTCCGTTGGTGGACTCTCAAAATC 136
QY 445 CCGACGACGAGAACGCTGATTCGATATGGGTGAAATCTGATGAAGAGAAACGTAA 504
Db 135 CCTCAACAAC--CTCAACAATAGTAATTCATTAACGAAACGCGGAAACCGGAAG 79
QY 505 CGACCTTCAACGATAGTATAGACATAAGCAGCGCGCTTTTCAGCTTC 554
Db 78 CGCCTTGAATTCGATTCGCGATATTAGTGAAGACGACCCCGTTTC 29

RESULT 14
BQ865644
LOCUS
DEFINITION BQ865644.YG.ab1 QG-ABCD1 lettuce salinas Lactuca sativa cDNA clone
Q865644, mRNA sequence.
ACCESSION BQ865644
VERSION BQ865644.1 GI:22251099
KEYWORDS EST
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
REFERENCE 1 (bases 1 to 747)
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison,
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
TITLE Lettuce and sunflower ESTs from the Compositae Genome Project
JOURNAL Unpublished
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [michelmore@ucdavis.edu]
singletons, see http://cgdb.ucdavis.edu/ for details.
Plate: Q865 row: H column: 15.
Location/Qualifiers
1. .747

FEATURES
source
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"

```

/db xref="taxon:4236"
/cclone="QC5H15"
/lab host="E.coli"
/cclone_lib="QC_ABCDI lettuce salinas"
/notes="Vector: pBRCN4S1A1B; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/TAG Lib=QC ABCDI lettuce salinas
TAG TISSUE=flowers pre-fertilized
TAG SEQ=GCCTGACGGG"
BASE COUNT 199 a 179 c 199 g 168 t 2 others
ORIGIN
Query Match 13.7%; Score 146.4; DB 13; Length 747;
Best Local Similarity 70.4%; Pred. No. 8.2e-35;
Matches 195; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

2Y 224 CTAAGACCTCACAGAGTTGAAGGAGAGGAGAGGATACGATGCTGCGACGT 283
Db 338 CTAAGACCTCACAGAGTTGAAGGAGAGGAGAGGATACGATGCTGCGACGT 397
2Y 284 GTCCGGCTAGATTTTCAATTAACTCGAGAGTTAGTTCACAAATCCGACGGGAGACGA 343
Db 398 GCGCTGCTAGAGTCTTCAGTTTACCTCGGAGTCTCGTTCACAGTACAGCGGAGACGA 457
2Y 344 TTGCGTGTGTTGGAGAACCTGACCGCGGATATACGCCACCGGATACGGGAACGG 403
Db 458 TTGCGTGTGTTGGAGAACCTGACCGCGGATATACGCCACCGGATACGGGAACGG 517
2Y 404 TTCCCGCATCGCATGTCGTTTACCGAACCTTAAATCCGACGACGACGACGCTG 463
Db 518 TTCCCGCATCGCATGTCGTTTACCGAACCTTAAATCCGACGACGACGACGCTG 577
2Y 464 ATTCTGATGCGTGAAATCTGATGAGAGAAACG 500
Db 578 GCAAGGAATGGAGATGCTAGGAGAGAGGAGAG 614

```

```

RESULT 15
3F070046
LOCUS
DEFINITION
BF070046 649 bp mRNA linear EST 06-DEC-2001
st63b02.y1 Gm-cl053 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl053-1227 5' similar to TR:064647 064647 PUTATIVE DNA BINDING
PROTEIN ; mRNA sequence.
BF070046
BF070046.1 GI:10847341
EST.
Glycine max (soybean)
Glycine max
SOURCE
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 649)
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ratter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800

```

```

Fax: 314 286 1810
Email: estowatson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
Insert Length: 1706 Std Error: 0.00
High quality sequence stop: 415.
Location/Qualifiers
1. 649
/organism="Glycine max"
/mol type="mRNA"
/db xref="taxon:3847"
/cclone="GENOME SYSTEMS CLONE ID: Gm-cl053-1227"
/tissue_type="Whole seedling, 3 week old, greenhouse
grown"
/lab host="DH10B"
/cclone_lib="Gm-cl053"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The Harsoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from whole seedlings of 3
week old greenhouse grown plants. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site and a 3'
anchor. EcoRI adapters were ligated to the blunt-ended
cDNA fragments followed by XhoI digestion. The cDNA
fragments were directionally cloned into the EcoRI-XhoI
restriction site of the pBluescript vector. The ligated
cDNA fragments were transformed into DH10B host cells
(GibcoBRL). This library was constructed in cooperation
with Dr. Paul Keim's laboratory at Northern Arizona
University."
BASE COUNT 158 a 162 c 170 g 159 t
ORIGIN
Query Match 13.6%; Score 145.8; DB 10; Length 649;
Best Local Similarity 66.2%; Pred. No. 1.2e-34;
Matches 210; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

```

2Y 136 GAGCCCAAGCGGAGCGGTGATGCGTGTTCATATCTTTAGTCCACCGTCTCG 195
Db 329 GTGCACGTGCGCGCGGGATGCTATATGCTCATGTCACAGGCTCTGCCCAAGCC 388
2Y 196 ACAGGACCACTTGAAGAGAGCTTCGACTAAGACCGTCACACGAAGTTGAAGAGA 255
Db 389 CAGCGCCACCAACAAAGCGGGCTCCACAAAGACCGCCACACCAAGTAGAGGGCGA 448
2Y 256 GCGAGAGAGATAGGATGCTGCCAGTGTGCGGTAGGATTTTCAATTAATCGAGAG 315
Db 449 GGCAGAAGGATTCGAATGCCCGCCACGTCGCGGAGGATCTTTTCAGCTGACCGAGAG 508
2Y 316 TTAGGTACAAATCCGACCGGAGAGATTCGTTGTTGGAGACGCTAGCGCGCG 375
Db 509 CTGCGTCAATAATCCGACCGGAGAACATCCCGTGGCTCCTCGAGCAGCGCGCGC 568
2Y 376 ATTATAGCGCCACCGGTACGGGAACGGTTCCCGGCATCGCCATGTCGGTTAACGGAAC 435
Db 569 ATTATTCGACGACCGGCAACGCGACAGGTTCGGCCATCCGATGTTGCGCATTTGAACG 628
2Y 436 TTAATAATCCGACGAC 452
Db 629 TAAAGATTTCGACCAC 645

```

Search completed: February 2, 2004, 02:56:33
Job time : 2486.95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

run on: February 2, 2004, 01:11:01; Search time 4063.8 seconds
(without alignments)
10781.591 Million cell updates/sec

title: US-09-938-842A-1034

effect score: 1071

sequence: 1 atggcgacaattcagaagct.....cacgggtcatcgaaaccactga 1071

scoring table: OLIGO-NUC
Gapop 60.0, Gapext 60.0

searched: 2888711 seqs, 20454813386 residues

ord size: 0

total number of hits satisfying chosen parameters: 5777422

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: listing first 45 summaries

atabase: GenEmbl:

```

1: gb.ba:*
2: gb.hg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.or:*
21: em.ov:*
22: em.pat:*
23: em.ph:*
24: em.pl:*
25: em.ro:*
26: em.sts:*
27: em.un:*
28: em.vi:*
29: em.hg.hum:*
30: em.hg.inv:*
31: em.hg.other:*
32: em.hg.mus:*
33: em.hg.pln:*
34: em.hg.pln:*
35: em.hg.rod:*
36: em.hg.mam:*
37: em.hg.vit:*
38: em.sy:*
39: em.hg.hum:*
40: em.hg.mus:*
41: em.hg.other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1071	100.0	1071	6	AX506339	Sequence
2	1071	100.0	1071	8	AF370606	Arabidops
3	1071	100.0	1102	8	AY113981	Arabidops
4	1071	100.0	1660	8	AY056214	Arabidops
5	1071	100.0	91854	8	AC003680	Arabidops
6	25	2.3	134151	8	AC006922	Arabidops
7	23	2.1	1206	8	BT008493	Arabidops
8	23	2.1	1479	8	AY081344	Arabidops
9	23	2.1	85961	8	AC079131	Arabidops
10	23	2.1	103637	8	AC079604	Arabidops
11	22	2.1	2134	3	AF352244	Arabidops
12	22	2.1	149452	2	AC124912	Arabidops
13	22	2.1	170737	2	AC125510	Arabidops
14	21	2.0	158514	2	AC051840	Arabidops
15	21	2.0	160796	9	AL499604	Human DNA
16	21	2.0	166983	9	AC002541	Human DNA
17	20	1.9	1099	10	AF176350	Rattus no
18	20	1.9	1233	10	AF218846	Rattus no
19	20	1.9	1305	10	AF383875	Rattus no
20	20	1.9	2709	10	AF286721	Rattus no
21	20	1.9	2926	10	AF286722	Rattus no
22	20	1.9	75493	2	AC015877	Homo sapi
23	20	1.9	94810	6	AX695407	Sequence
24	20	1.9	99007	2	AC142506	Medicago
25	20	1.9	110000	2	AC121458	Continuation (2 of
26	20	1.9	110000	2	AC121458	Continuation (3 of
27	20	1.9	110000	2	LMFLCHR36	Continuation (22 of
28	20	1.9	124347	9	AC010072	Continuation (22 of
29	20	1.9	131701	9	AC110583	Homo sapi
30	20	1.9	136064	10	AL928992	Mouse DNA
31	20	1.9	136150	8	AF002485	Oryza sat
32	20	1.9	136579	9	HS998020	Human DNA
33	20	1.9	141036	2	AC023807	Human DNA
34	20	1.9	141227	2	AC137836	Medicago
35	20	1.9	144820	2	AC099556	Trypanoso
36	20	1.9	146301	2	AC021414	Homo sapi
37	20	1.9	149042	2	AC016410	Homo sapi
38	20	1.9	152990	2	AC130432	Homo sapi
39	20	1.9	160359	9	AC007262	Homo sapi
40	20	1.9	160846	9	AC124306	Homo sapi
41	20	1.9	169337	2	AC123755	Mus muscu
42	20	1.9	170867	9	AC068295	Homo sapi
43	20	1.9	172159	2	AF005492	Oryza sat
44	20	1.9	178461	9	AL357075	Human DNA
45	20	1.9	179155	10	AL772401	Mouse DNA

ALIGNMENTS

RESULT 1
AX506339
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

AX506339
Sequence 1034 from Patent WO0216655.
AX506339
AX506339.1 GI:23387576

Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1
Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
Stress-regulated genes of plants, transgenic plants containing

linear PAT 27-SEP-2002

same, and methods of use
Patent: WO 0216555-A 1034 28-FEB-2002;
The Scripps Research Institute (US); Syngenta Participations AG
(CH)

JOURNAL
FEATURES
source
Location/Qualifiers
1. 1071
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"

BASE COUNT 287 a 288 c 256 g 240 t

ORIGIN

Query Match 100.0%; Score 1071; DB 6; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGGCAATTCAGAGCTTGAAGAGTTGCGAGGAAAGATCAAACTTAAGAGCGTT 60
DB 1 ATGCGGCAATTCAGAGCTTGAAGAGTTGCGAGGAAAGATCAAACTTAAGAGCGTT 60
QY 61 GATCTAACCATCATCAACGGCTCGAAGAGCTGGAAGCTTCAAGACTTCCCAAGTAAT 120
DB 61 GATCTAACCATCATCAACGGCTCGAAGAGCTGGAAGCTTCCCAAGTAAT 120
QY 121 CCCACAGTGTCTCGAGCCCAAGCGGAGCGGTGATGCGTCTTCAATGCTTTA 180
DB 121 CCCACAGTGTCTCGAGCCCAAGCGGAGCGGTGATGCGTCTTCAATGCTTTA 180
QY 181 GCTCCACCTCTTCGACAGGACACCATTCAGAGAGCTTCGACTAAGACGTCACAG 240
DB 181 GCTCCACCTCTTCGACAGGACACCATTCAGAGAGCTTCGACTAAGACGTCACAG 240
QY 241 AAGGTTGAAGAGGAGGAGAGATACGGATGCTGCCACGTTGCGGCTAGGATTTT 300
DB 241 AAGGTTGAAGAGGAGGAGAGATACGGATGCTGCCACGTTGCGGCTAGGATTTT 300
QY 301 CAATTAAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 360
DB 301 CAATTAAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 360
QY 361 AAGCTGAGCGCGGATTTATAGCGGCTACGGGAGCGGTTCGCGCATCGCGATG 420
DB 361 AAGCTGAGCGCGGATTTATAGCGGCTACGGGAGCGGTTCGCGCATCGCGATG 420
QY 421 TCGGTTAAGCGAACCTTAAATATCCGAGCGAGCGAGCTGATCTGATGAGTGA 480
DB 421 TCGGTTAAGCGAACCTTAAATATCCGAGCGAGCGAGCTGATCTGATGAGTGA 480
QY 481 AATCTGATGAAGAGAAACGTTAAAGAGCTTCTAAGAGTGTATATAGATAGCGAC 540
DB 481 AATCTGATGAAGAGAAACGTTAAAGAGCTTCTAAGAGTGTATATAGATAGCGAC 540
QY 541 GCGGTTTACGTTCTCGGTTTGTAGTCCAAATGCGAGCGAGCAACGATCCAACTCG 600
DB 541 GCGGTTTACGTTCTCGGTTTGTAGTCCAAATGCGAGCGAGCAACGATCCAACTCG 600
QY 601 CAAGCTCTGCATCATCATCTGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 660
DB 601 CAAGCTCTGCATCATCATCTGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 660
QY 661 TGGGCTTATTCATCAACGCAATGATTCGAGCGTGGAGCTTTCTTTTGTATTCACAA 720
DB 661 TGGGCTTATTCATCAACGCAATGATTCGAGCGTGGAGCTTTCTTTTGTATTCACAA 720
QY 721 ATCGTCTGCTCGGATAGCTGATTTATAGCTTTTCCCGCGCGCTGCTTCGCGG 780
DB 721 ATCGTCTGCTCGGATAGCTGATTTATAGCTTTTCCCGCGCGCTGCTTCGCGG 780
QY 781 TCGTCTTACGTCGCCGCTGTTCACAGGCTTCCACGATGGCTAGACCACTCTTTTACAA 840
DB 781 TCGTCTTACGTCGCCGCTGTTCACAGGCTTCCACGATGGCTAGACCACTCTTTTACAA 840
QY 841 GTTGTTCAGAGAGCGGCTTTGTATCGGTTTCAGAGTTAGCGGTTTCGAATTTATCAAGA 900

DB 841 GTTGTTCAGAGAGCGGCTTTGTATCGGTTTCAGAGTTAGCGGTTTCGAATTTATCAAGA 900
QY 901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 960
DB 901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 960
QY 961 ATTGCAACCAACAGAGCGCACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAACAA 1020
DB 961 ATTGCAACCAACAGAGCGCACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAACAA 1020
QY 1021 GAGCTTACCACTTCATGAGCACCACAGCAGCAGCTGATCGAACCACTCA 1071
DB 1021 GAGCTTACCACTTCATGAGCACCACAGCAGCAGCTGATCGAACCACTCA 1071

RESULT 2
LOCUS AF370606 1071 bp mRNA linear PLN 30-APR-2001
DEFINITION Arabidopsis thaliana putative PCF2-like DNA binding protein
ACCESSION AF370606
VERSION AF370606.1 GI:13877694
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (Bases 1 to 1071)
AUTHORS Lan, B.; Southwick, A.; Karlin-Neuman, G.; Nguyen, M.; Miranda, M.;
Pal, C. J.; Bowser, L.; Jones, T.; Ban, J.; Chen, H.; Cheuk, R.;
Chung, M. K.; Kim, C.; Lin, J.; Liu, S. X.; Pham, P. K.; Sakano, H.;
Shinn, P.; Yamada, K.; Becker, J.; Theologis, A. and Davis, R. W.
Direct Submission
Submitted (17-APR-2001) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT e-mail for correspondence: arab@sequence.stanford.edu
This clone was isolated by RT-PCR.
FEATURES
source
1. 1071
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="II"
/clone="CS15"
/note="ecotype: Columbia"
gene
1. 1071
/gene="A:2945680; F17K2.21"
CDS
1. 1071
/gene="A:2945680; F17K2.21"
/codon_start=1
/product="putative PCF2-like DNA binding protein"
/protein_id="AAK4325.1"
/db_xref="GI:13877695"
translations="MARIQKLEEVAGKQOTLRAVLTIIIVRVNVEIRPFOVNPVLS
LEPAKFEVPEFMSAPSSPPPLKRASTKDRKFKVEGRGRIRMPATCAARIFQL
TRELGKSGDETIRILLENAEPAIATATGTVPALAMSVNGTKLPTITNADSDMG
NLMKTRKPSNSEYIDIDAVSAGSLAPIATTTIOPQALASTVAQOLLPOQMY
PMWAI PSNAMIPTVGAFILIPQAGSNQPIILAPPAASPSYVA/VQOASTWARD
PELOVPSGPFVSVDVSGNSLRATSWAPSSSGVTGSSSSATTTTTLRDFSL
EYKQELQGFMTTPARSNH"

BASE COUNT 287 a 288 c 256 g 240 t

ORIGIN

Query Match 100.0%; Score 1071; DB 8; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGGCAATTCAGAGCTTGAAGAGTTGCGAGGAAAGATCAAACTTAAGAGCGTT 60
DB 1 ATGCGGCAATTCAGAGCTTGAAGAGTTGCGAGGAAAGATCAAACTTAAGAGCGTT 60

	Db	643	GCGCTTTTCAGCTTCTCCGGTTTAGCTTCAATTTGACACGACGACGACAAGATCCAACTCCG	702
Qy	601	CAGGCTCTGGCATCATCACCATGTGGCTCAGCAACTTCTGCCGCGAAGGAATGTATCCGATG	660	
Db	703	CAGGCTCTGGCATCATCACCATGTGGCTCAGCAACTTCTGCCGCGAAGGAATGTATCCGATG	762	
Qy	661	TGGGCTATTTCATCAAACGCATGATTCCGACGGTCGGAGCTTTCTTTCTTGATTTCCACAA	720	
Db	763	TGGGCTATTTCATCAAACGCATGATTCCGACGGTCGGAGCTTTCTTTCTTGATTTCCACAA	822	
Qy	721	ATCGCTGTGTCGGTCGATCAGCCTCAGTTATTAGCTTTTTCCGCGCGCGCTGCTTCGCCG	780	
Db	823	ATCGCTGTGTCGGTCGATCAGCCTCAGTTATTAGCTTTTTCCGCGCGCGCTGCTTCGCCG	882	
Qy	781	TGCTCTTTACGTCGCGCGCTGTTCACACAGGCTTCCACGATGGCTAGACCACTCTCTTTACAA	840	
Db	883	TGCTCTTTACGTCGCGCGCTGTTCACACAGGCTTCCACGATGGCTAGACCACTCTCTTTACAA	942	
Qy	841	GTTGTTCCAAGCAGCGGCTTTGTAATCGTTTTCAGAGCTTTCAGAGCTTTCAGAGCTTTCAGAG	900	
Db	943	GTTGTTCCAAGCAGCGGCTTTGTAATCGTTTTCAGAGCTTTCAGAGCTTTCAGAGCTTTCAGAG	1000	
Qy	901	GCGAGCTCGGTTATTTGGCTTCGAGCTCAAGCTCAGGCGTAAACACCGGTAGTTTCATCGTCA	960	
Db	1003	GCGAGCTCGGTTATTTGGCTTCGAGCTCAAGCTCAGGCGTAAACACCGGTAGTTTCATCGTCA	1060	
Qy	961	ATTGCAACAACAAGCAGCACGCTCAGAGCTTTCCTCTAGAGATATACGAGAACA	1020	
Db	1063	ATTGCAACAACAAGCAGCACGCTCAGAGCTTTCCTCTAGAGATATACGAGAACA	1120	
Qy	1021	GAGCTTCAACCAAGTTTCATGAGCACCACACAGCAGGTCATCGAACCACTGA	1071	
Db	1123	GAGCTTCAACCAAGTTTCATGAGCACCACACAGCAGGTCATCGAACCACTGA	1173	
RESULT 5				
LOCUS	AC003680/c			
DEFINITION	Arabidopsis thaliana chromosome 2 BAC F1/K2 genomic sequence,			
ACCESSION	AC003680			
VERSION	AC003680.3			
SOURCE	HTG.			
ORGANISM	Arabidopsis thaliana (thale cress)			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi			
	1 (bases 1 to 91854)			
	Lin,X., Kaul,S., Town,C.D., Benito,M.I., Creasy,T.H., Haas,B.J.,			
	Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,			
	Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.			
	Arabidopsis thaliana chromosome 2 BAC F1/K2 genomic sequence			
	Unpublished			
	2 (bases 1 to 91854)			
	Lin,X.			
	Direct Submission			
	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712			
	Medical Center Dr., Rockville, MD 20850, USA			
	3 (bases 1 to 91854)			
	Town,C.D. and Kaul,S.			
	Direct Submission			
	Submitted (27-FEB-2002) The Institute for Genomic Research, 9712			
	Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org			
	On Apr 18, 2002 this sequence version replaced gi:6598396.			
	Address all correspondence to:at@tigr.org			
	BAC clone F1/K2 is from Arabidopsis thaliana chromosome 2 and is			
	near the molecular marker(s) FUS.			
	The orientation of the sequence is from SP6 to T7 end of the BAC			
	clone.			
	Genes were identified by a combination of several methods: Gene			
	prediction programs including GENSCAN. (Chris Burge.			

<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkES (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerW, see Mihaela Perce, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihaela Perce and Steven Salzberg, contact mperte@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding TRMs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Location/Qualifiers
1. 91854
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C1C02E07"
/clones="F17K2"
complement(1..1281)
/notes="overlap with BAC clone F4123 (AC002387.1..1281)."
57..747
/gene="At2g45480"
/notes="synonym: F17K2.27"
join(<57..242,335..>747)
/gene="At2g45480"
join(57..242,335..747)
/gene="At2g45480"
/notes="Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAPI8607.3"
/db_xref="GI:20197053"
/translations="DESKPSVQKFFPEVSDKLEAAKFSNRKNDIARSRWKNNV NGSLFGHIGRSPVTLQRCFCFLQGVETDNEPGRCRIDGKRWKCKDLVSGQYCD KHRHGRKKHPVDVTSNHNAGFSLPVTAVRVPCKDGDQKHVSVMGITLPR VDEKTSSTCTTTITDTRALGEDDDEEYLSLSPGV"
complement(1002..2388)
/gene="At2g45490"
/notes="synonym: F17K2.2; contains a protein kinase domain profile (PDOC00100)"
complement(join<1002..1112,1208..1280,1366..1580,1655..1750,1886..2128,2260..>2388)
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,1655..1750,1886..2128,2260..2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"
/translations="MSKKSTESDAGNTEKQWLSADPEIGRPLGKGFGRVYLAREAKS KYIVALKVIFKEQIKYHQHREMEICTSLRHENILRLFGWFDHNERIFLILEYA HGGLYGLVKONGHLTEQQAQVIASLSQALYACGKCVIHEDIKPENLLDHEGLK IADFGVSSNRKRTMCGTLDYLAPEMVENRDHYADVNTWLGILCVFLYGNPPE AESQKDTFRLLKIDLSFLPLTPNVSEAKNLISQLLVKDPKRLSIEKIMQHPWTRN ADPKGVCASIDI"
complement(3414..5980)
/gene="At2g45500"
/notes="synonym: F17K2.3; predicted by genscan"
complement(join<3414..3616,3639..3825,3916..3965,4032..4215,4582..4712,4769..4864,4948..5188,5352..5374,5506..5585,5670..>5980))
/gene="At2g45500"

CDS

complement(join(3414..3616,3639..3825,3916..3965,4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,5506..5585,5670..5980))
/gene="At2g45500"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC06152.1"
/db_xref="GI:2979543"
/translations="MSFLRGIIDFSILNEESKDPSSVSSSTSSSSMNGIDGPVPT NERAYKLGYFDLAKBEIAKGVRAEWHLDHALLHYRNAORIMNEATSTPSPSYIS SSEKVRYSREKTSNNQVSEBLOALGVMSNKETVAVPSASVSTSPSRKTL SOKTPVAGGVATPRNEKDAASPVKVKGSGNYDDKLVEINTIIVDRSPSVKDDV AGLNCAKALLBMLVLPKARDLFTGLRRPARVTSLLGLLFGPPGKNTLAKVAS ESQATFNVSASSLTSKWDISIMSTRSTSEARSLKSEFLIQFDGVTNPDVLVII IGANKQELDDAVLRVLRVYVLPEDSNVRKLLFKLKQPHSLSDGDIKRYKE TEGKLYRLCIKRRFISQVTDKRYGSDLOALCEEAAMPFIREIGANILITQANKVLN FVSQINVEVCLSLKGLINRCVHTGKIST"
6429..8288
/gene="At2g45510"
/notes="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"
/translations="MEITGIAITVATTIVLCFTIYLMIRIFTGSKRNDKRYAPVH ATVPDLLHSDLYDYTEIARSKPTVPLSPGQSILTADPRNVHILKTRFDNYSK GHSSRENMDLLGHGIPAVDGEKWRQKLSSEFSTRVLDRPSCSVFERNASKLVGF VSEFALSGKAFDAQDQLMRCLTDSIEKVGFGVELKLDGSKEGQGFMEAFDEGNVAT SSRFIDELMLKFFNIGSOSKUKLSTATIDKVEYSLITTKKELAKEQNTVVREDIL SRFVSEKQDNNNDKYLDIILNPMIAGKDTTALLSNFLMCKNPLVQEKIYOE LPDGRVSKGDNITYIAYAMGRYIWGQDAEFPKPERWLKGLFGPSPGPFISFHA GPRLCKDFAYRQMKIVSMALLHFFRFKWDENSKVYKRMILTHVDGGLHLCAPR TST"

gene

mRNA

CDS

gene

complement(8412..9632)
/gene="At2g45520"
/notes="synonym: F17K2.5; supported by full length cDNA:
Ceres:8220"

mRNA

complement(join<8412..8678,8768..8903,8987..9120,9289..9356,9459..>9632))
/gene="At2g45520"

CDS

complement(join(8553..8678,8768..8903,8987..9120,9289..9356,9459..9612))
/gene="At2g45520"
/codon_start=1
/product="expressed protein"
/protein_id="AAC06155.1"
/db_xref="GI:2979546"

/translations="MGKSKKREKRYLAHGGPARLPPPPRSKQDDVPSTIRLMN YTSFSDRSKQVVEKKEKKAQVVPATSDSDSVVEKKKKRKNQMTDLRFEN ELAIDRSKRRKKKWEAKQKKGKGTEDTLRENFPEHQIRFGDVGQAPLKLA VVPAKRSKLASQERLQAIDAYRSRGWTARPGVIPVMMQ"
9408..9427
/rpt_family="(CA)n"
complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"

repeat_region

repeat_region

gene

complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"

mRNA

complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"

CDS

complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"

complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"

complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"

complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"

```
/db_xref="GI:2979545"
/translation="WQSSSSGGDLLEKQQQQQKDKPQKQSSANESNHLITIVCN
GDSREELVQIPPEKVELSNGSHQCRVCLQDEBVLIELGQCRCGLAKAHS
CIDAFRRKGNQCSICQVAVNVPPTPTNNWWRIDPSYRQEREGCFSPW
VAFSILIGLMLDLVLSITLGVNSALFVNIIIGVIVLGLTALRLTLEFCVEMSLRA
VHRVQRAEANNFNIAVPPAL"
repeat_region 9988..10007
gene /rpt_family=" (CAG)n"
complement(10833..25573)
/gene="At2g45540"
/note="synonym: F17K2.7"
complement(join(<10833..11272,11438..11560,11636..12378,
12454..12937,13023..13119,13254..13374,13464..13548,
13645..13794,13957..14073,14178..14264,14350..14418,
14520..14576,14730..14813,14926..14995,15396..15538,
15768..16244,16322..16465,16545..17093,17237..17431,
17612..17896,18015..18613,18944..19249,19633..19714,
19972..20048,20151..20264,20429..20533,20796..20930,
Query Match 100.0%; Score 1071; DB 8; Length 91854;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 ATGGCGCAATTCAGAGCTTGAAGATTCGAGCAAGATCAAACTCTAAGAGCGTT 60
b 75131 ATGGCGCAATTCAGAGCTTGAAGATTCGAGCAAGATCAAACTCTAAGAGCGTT 75072
Y 61 GATCTAACCATCATCAACGGGTGAGAACTCGAACTTCAAGACCTTCCAGTAAAT 120
b 75072 GATCTAACCATCATCAACGGGTGAGAACTTCAAGACCTTCCAGTAAAT 75012
Y 121 CCCAGTGAATCTCGAGCCCAAGCGGAGCGGTGATCGCGTGTCTTCAATGCTTTA 180
b 75011 CCCAGTGAATCTCGAGCCCAAGCGGAGCGGTGATCGCGTGTCTTCAATGCTTTA 74952
Y 181 GCTCACCGTCTTCGACAGGACCACTTGAAGAGCTTGCATTAAGACCGTCAACG 240
b 74951 GCTCACCGTCTTCGACAGGACCACTTGAAGAGCTTGCATTAAGACCGTCAACG 74892
Y 241 AAGTTGAAGGAAGGAGGAGGATAGATCGCTCCAGCTGCGGCTAGGATTTT 300
b 74891 AAGTTGAAGGAAGGAGGAGGATAGATCGCTCCAGCTGCGGCTAGGATTTT 74832
Y 301 CAATTAATCTGAGATTTAGTTCACAAATCCGACGGCGAAGATTCGGTGTGAG 360
b 74831 CAATTAATCTGAGATTTAGTTCACAAATCCGACGGCGAAGATTCGGTGTGAG 74772
Y 361 AACGCTGAGCGGCGATTTATAGCCGACGGTACGGAAACGTTCCGCGCATCGCATG 420
b 74771 AACGCTGAGCGGCGATTTATAGCCGACGGTACGGAAACGTTCCGCGCATCGCATG 74712
Y 421 TCGGTTAACGAACCTTAAATAATCCGACGACGACGACGCTGATTCGATATGGGTGA 480
b 74711 TCGGTTAACGAACCTTAAATAATCCGACGACGACGACGCTGATTCGATATGGGTGA 74652
Y 481 AATCTGATGAGAGAACGTAACGACCTTCAACAGTGTGATATAGACATAAGCGAC 540
b 74651 AATCTGATGAGAGAACGTAACGACCTTCAACAGTGTGATATAGACATAAGCGAC 74592
Y 541 GCCGTTTCAGCTTCTCGGTTTACGTTTCAATTCGACGACGACGATCCAACTCCG 600
b 74591 GCCGTTTCAGCTTCTCGGTTTACGTTTCAATTCGACGACGACGATCCAACTCCG 74532
Y 601 CAAGCTTGGCATCATCACTGTGGCTCAGCAATTCGCGGCAAGGAATGTATCGATG 660
b 74531 CAAGCTTGGCATCATCACTGTGGCTCAGCAATTCGCGGCAAGGAATGTATCGATG 74472
Y 661 TGGGCTATTCATCAACAGCAATGATTCGACGCTCGAGCTTCTTCTGATTCACAA 720
b 74471 TGGGCTATTCATCAACAGCAATGATTCGACGCTCGAGCTTCTTCTGATTCACAA 74412
Y 721 ATGCTGCTCGTTCGATTCAGCTTCAGCTTATAGCTTTTCCGCGCGCTGCTCGCGG 780
|||||
```

```
Db 74411 ATCGCTGGTCCGTCGAAATCAGGCTCAGTTATTAGCTTTTCCGCGCGCTGCTTCGCGG 74352
Qy 781 TCGTCTTAGTTCGCGCTGCTTCAACAGGCTTCCAGATGGCTAGACCACTCTTACAA 840
Db 74351 TCGTCTTAGTTCGCGCTGCTTCAACAGGCTTCCAGATGGCTAGACCACTCTTACAA 74292
Qy 841 GTTGTTCAGAGCGCGGCTTGTATCCGTTTCAGACGTTAGCGGTTCAATTTATCAAGA 900
Db 74291 GTTGTTCAGAGCGCGGCTTGTATCCGTTTCAGACGTTAGCGGTTCAATTTATCAAGA 74232
Qy 901 GCGACGCTCGGTTATGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGTAGTTCATCGTCA 960
Db 74231 GCGACGCTCGGTTATGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGTAGTTCATCGTCA 74172
Qy 961 ATTGCAACAAACAGCGGACACAGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACAA 1020
Db 74171 ATTGCAACAAACAGCGGACACAGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACAA 74112
Qy 1021 GAGCTTCACGAGTTATGAGACCAACAGCAGCAGCTCATCGAACCACTGA 1071
Db 74111 GAGCTTCACGAGTTATGAGACCAACAGCAGCAGCTCATCGAACCACTGA 74061

RESULT 6
LOCUS AC006922
DEFINITION Arabidopsis thaliana chromosome 2 clone T118 map 96825, complete
sequence.
ACCESSION AC006922
VERSION AC006922.7 GI:20197941
KEYWORDS HTG
SOURCE Arabidopsis thaliana (chale cross)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie.
REFERENCE 1 (bases 1 to 134151)
AUTHORS Lin X., Kaul S., Shea, T.P., Fujii, C.Y., Shen, M., VanAken, S.E.,
Barnstead, M.E., Mason, T.M., Bowman, C.L., Renning, C.M.,
Benito, M., Carrera, A.J., Creasy, T.H., Buell, C.R., Town, C.D.,
Niernan, W.C., Fraser, C.M. and Venter, J.C.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 134151)
AUTHORS Lin X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 134151)
AUTHORS Town, C.D. and Kaul, S.
TITLE Direct Submission
JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdtown@igir.org
COMMENT On Apr 18, 2002 this sequence version replaced gi:6598638.
On Oct 8, 1997 this sequence version replaced gi:243866.
We have determined that TAC YUP8H12 is chimaeric, and is comprised
of two distinct genomic EcoRI fragments from chromosome 1. This
submission contains the sequence from the EcoRI site at position 1
(right end) to position 181918 of our previous Phase II
submission YUP8H12 accession number AC000098. This fragment maps
at the bottom of the right arm of the chromosome between the ATHATPAS
and m532 markers.
FEATURES
source
1..134151
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="96825"
/clones="T118"
misc_feature 1..2183
```


/note="overlap with BAC clone F13K3
 (AC006282:90190..92376)."
 complement(228..251)
 /rpt_family="AT_rich"
 1600..1690
 /rpt_family="AT_rich"
 3774..3801
 /rpt_family="AT_rich"
 4362..4397
 /rpt_family="AT_rich"
 5421..5442
 /rpt_family="AT_rich"
 5481..5527
 /rpt_family="AT_rich"
 6605..7737
 /gene="At2g36830"
 /note="synonym: T1J8.1; supported by full length cDNA:
 Cere9:36633"
 join(6605..7052,7144..7737)
 /gene="At2g36830"
 6640..6996
 /note="molecular marker ve017"
 join(6669..7052,7144..7515)
 /gene="At2g36830"
 /codon_start=1
 /product="putative aquaporin (tonoplast intrinsic protein
 gamma)"
 /protein_id="AAD31569.1"
 /db_xref="GI:4883600"
 /translation="MPTRTAIGRPDTRPDAKALAFISTLIPVAVSGSGMAF
 NKLTEWATPSGVAAVAFAFGLFVAVSGVANGISGVNPAVTFGAFIGNITLLR
 GILVIAQLGLGVACILKFAFGGLAVPAFGLSAGVGLNPAFVIVMTFGLVTVY
 ATAIDPKNGLTIAPITGIPTVIGANTLAGAFSGASMPAVAFGPAVSWTWNHW
 YWAGPLVGGGTLGLIYEVFFINTHEQLPTDY"
 7847..7891
 /rpt_family="AT_rich"
 7952..7981
 /rpt_family="(A)n"
 complement(10803..12597)
 /gene="At2g36835"
 /note="synonym: T1J8.24; supported by full length cDNA:
 Ceres: 97088"
 complement(join(<10803..11068,11697..11765,12150..12260,
 12343..12597))
 /gene="At2g36835"
 complement(join(11018..11068,11697..11765,12150..12360,
 12343..12486))
 /gene="At2g36835"
 /codon_start=1
 /product="Expressed protein"
 /protein_id="AAM15323.1"
 /db_xref="GI:20197944"
 /translation="MAKGGATLAEDAPRWKSSGREVPKISRPSVLISQNPETDIYAI
 SVMKHPNPGVGFAMEAVLSEAGSECVPGQVPTLLGLLVKVPVEVDLKEPVGKE
 LKMLGFMDNAVELMKNKSFIDR"
 complement(12752..14805)
 /gene="At2g36840"
 /note="synonym: T1J8.2; predicted by genefinder; supported
 by full length cDNA: Ceres:108427"
 complement(join(<12752..13150,13249..13533,13650..13971,
 14054..14159,14237..14464,14576..14805))
 /gene="At2g36840"
 complement(join(13013..13150,13249..13533,13650..13971,
 14054..14159,14237..14464,14576..14729))
 /gene="At2g36840"
 /codon_start=1
 /product="expressed protein"
 /protein_id="AAD31570.1"
 /db_xref="GI:4883601"
 /translation="MGILSDVVIIISQSEKGDPSVITNCPTDKGLCDLCLRIILFF
 GLNIVGSDVTDGKCYLVFWIGKNTRWNLKMLVPAASFAGISRCYISDS
 ESQPKLPDLFLKLACSDRTGLLYDTEVLYKLEINIEKVKISTTPDGKNDLPFVT

gene

mRNA

CDS

gene

tRNA

repeat_region

gene

mRNA

CDS

repeat_region

repeat_region

repeat_region

repeat_region

DTRELLGTAKREVEYELRDAIGDSMISYDIELVGEITACSTSSSSVAETLFSDDVS
 GEHSSGLTNSVSIADNSLSSAHILHITCODHGLLYIMRTFKDFNIOYGRF
 TIKKNCRIIDLFTVQSDRKILDSKLNALITRLRAELOQLPLRVMMNRGPDTELV
 TNPVLSGKGRPOPHDIALALKKIDTCIFSAEIGRHVGTGDRVEVVKVLNEEDSLP
 IPRSLIEEVNKLWGWE"
 complement(15868..27297)
 /gene="At2g36850"
 /note="synonym: T1J8.3"
 complement(join(<15868..16473,16628..16678,16797..16916,
 16997..17099,17634..17773,17867..18035,18602..18764,
 18879..18921,19083..19201,19620..19735,20017..20136,
 20301..20424,20823..20988,21079..21147,21292..21426,
 21519..21628,21711..21871,27202..27297))
 /gene="At2g36850"
 complement(join(16114..16473,16628..16678,16797..16916,
 16997..17099,17634..17773,17867..18035,18602..18764,
 18879..18921,19083..19201,19620..19735,20017..20136,
 20301..20424,20823..20988,21079..21147,21292..21426,
 21519..21628,21711..21871,27202..27297))
 /gene="At2g36850"
 /codon_start=1
 /product="putative glucan synthase"
 /protein_id="AAD31571.1"
 /db_xref="GI:4883602"
 /translation="MFTPTACFELSWMPTKTESFLSKPKRKYVFTQVKRLHLIT
 VKDAANVPKNLEARRLEFPTSLFMDNQAAPVAPVPSVETPYSTVLYSSSE
 LRSNEQGISILFYLOKIFDEVENFLERGRSESTQADLOASDIALSERFWVSIR
 QQTARVTRGMYTRRALMLQSLFERKGLVDASLTNMPRGFESSLEAQAADLKPT
 VSCQIYQOQKQKPEATDIGLLQRYEARLAVFTSHSDVNGDGGSGGKPEFYSK
 LKADIIHGKEEYISIKLPGDKLGEKPNQNHAIVFTGEATQITDNQDNLBSEA
 IKMNLLEFPHGKHGIRRTILGVREHFTGRVMHVHDPDFDRIPIFRGGISKAS
 RVINISDIIYAGMFNSTLAQNTIHHEDVGLNQLFEKGVAGNGEQLSRDVRI
 GOLDFRMMGSEYFTTVGVCTVMAFSGADLAIRVAKLSGNTALDAALNAQFLVQIG
 IFTAPVMVGFILLEGILLKAIKPSITMQFQICSVFFPSLGTTHFTGRTILLAGAKV
 VLLYIAWTVEDFDWVSWLMYKGVGVKGLSWNEEEOAHIOIQLRIETLILS
 LRFPFQYGVYKLDTRKNTSLALXGYSWVILVIVFLFKGVASITFTALIVATAM
 TDSIPDMFACVLGFIPTGWALLSLATWKLQVLRVLGLNETVREFGRIYDAAMGLIF
 SPIALSWFFPFIPTFQSRLLFNQAFSRGLSILLAGNANVET"
 complement(31341..31424)
 /gene="At2g36860"
 complement(31341..31424)
 /gene="At2g36860"
 /product="tRNA-Tyr"
 complement(32969..33006)
 /rpt_family="(TA)n"
 complement(33824..35833)
 /gene="At2g36870"
 /note="synonym: T1J8.5; similar to cellulase (xyloglucan
 endo transglycosylase) GI:311835 from (Tropaeolum majus);
 supported by full length cDNA: Ceres: 7831"
 complement(join(<33824..34414,35094..35311,35394..35494,
 35596..35833))
 /gene="At2g36870"
 complement(join(34048..34414,35094..35311,35394..35494,
 35596..35833))
 /gene="At2g36870"
 /codon_start=1
 /product="xyloglucan endotransglycosylase, putative"
 /protein_id="AAD31572.1"
 /db_xref="GI:4883603"
 /translation="MGNLSILSLIFHLAVLGGSSVNAVPPSPGWPSSKVGSLNFI
 KGRNLNGHQHQMDQNALITWDRSGSGFKSVKFRSGYFGANIKQPGTAVGIT
 SLYSNNEAHPGHDEVDIEFLGTTEGKYTIQTNVYIGSGDGKLIIGREKFLWFD
 PTQDFHAILWSPREIIFVDIDIPRRYPKKSATFFPLRPMWLGSNDASWATED
 GKADYKQPTAKYTNFKALGCTAYNSARCYPLSAPYRSGGLTROHQHARRVQT
 HSMVNYCKDYKEDHSLTPECWR"
 complement(34562..34600)
 /rpt_family="AT_rich"
 complement(36788..36911)
 /rpt_family="AT_rich"
 37494..37687
 /rpt_family="AT_rich"
 complement(37725..37745)

```

repeat_region 37801 37845 "AT_rich"
repeat_region 38111 38146 "AT_rich"
gene complement(40567..42581)

```

```

Query Match 2.3% Score 25; DB 8; Length 134151;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

2Y 222 GACTAAGACCGTCACACGAGGTT 246
Db 101694 GACTAAGACCGTCACACGAGGTT 101718

```

```

RESULT 7
BT008493 Arabidopsis thaliana At1g58100 gene, complete cds.
DEFINITION
ACCESSION BT008493
VERSION BT008493.1 GI:30725659
KEYWORDS
SOURCE
ORGANISM

```

```

Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1206)

```

```

AUTHORS
Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Dale,J.M., Hayashizaki,Y., Hsuan,V.W., Ishida,J., Jones,T.,
Kamuya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Arabidopsis ORF clones
Unpublished
2 (bases 1 to 1206)

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Dale,J.M., Hayashizaki,Y., Hsuan,V.W., Ishida,J., Jones,T.,
Kamuya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (15-MAY-2003) Salk Institute Genomic Analysis Laboratory
(SiGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

```

```

TITLE
JOURNAL
COMMENT
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamuya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

```

The Salk, Stanford, PGC (SSP) Consortium members constructed and sequenced the PUMI (ORF) clones using the RAFL cDNAs: Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Chan,M.M., Chang,C.M., Dale,J.M., Hsuan,V.W., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S., Davis,R.W., Theologis,A., and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs.

```

FEATURES
Location/Qualifiers
1..1206
/molecule="mRNA"

```

```

/db_xref="taxon:3702"
/chromosome="1"
/clone="U24662"
/note="This clone is in pUNI 51
ecotype: Columbia"
1..1206
/note="unknown protein"
/codon_start=1
/product="At1g58100"
/protein_id="AAP37852.1"
/db_xref="GI:30725660"
/translation="MDLSIRNNNTAAVATGGARQLVDASLSIVPRSTPESTL
ATTSSATATTTKRSYKDRHTKVDGRIRMPALCARVFTQLRELKHSQTHQYD
LQQAFAIVARTGTITPANFTSLVSRSSGSLSPSPKSVPLYGALGTHQYD
EQGGVFAAHTSPILGFLHQHQNQNDPVTIPEGENSKRYRKYVDLSKEN
DRAKONENKSESTSGPTAAFWAVAPSSAGNFTWMLPVPPTAGNOMESSNN
NTAAGHRAPMPWPFVNSAGGACGGAATHFMAGTGFSPMDQYRGSPQLQGSFLAQ
PQTONLGLSMPDPSNLGLMLAALNSAYSRGNANAEQANNAVEHKEKQSQSHDDDS
RENSNSSE"

```

```

BASE COUNT 328 a 283 c 326 g 269 t
ORIGIN

```

```

Query Match 2.1% Score 23; DB 8; Length 1206;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY 220 TCGACTAAGACCGTCACACGAA 242
Db 175 TCGACTAAGACCGTCACACGAA 197

```

```

RESULT 8
AY081344 Arabidopsis thaliana 1479 bp mRNA linear PLN 24-MAR-2002
LOCUS Arabidopsis thaliana auxin-induced basic helix-loop-helix
DEFINITION transcritpion factor, putative (At1g58100) mRNA, complete cds.
ACCESSION
VERSION AY081344.1 GI:19698994
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1479)

```

```

REFERENCE
AUTHORS
Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamuya,A.,
Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (26-FEB-2002) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
e-mail for correspondence: arab@sequence.stanford.edu

```

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamuya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES

Location/Qualifiers

1..1479
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="RAF109-64-H20 (R24662)"
/note="This clone is in pBluescript
ecotype: Columbia"
1..1479
/gene="Atig58100"
/note="synonym: T15M6.11"
43..1248
/gene="Atig58100"
/codon_start=1
/product="auxin-induced basic helix-loop-helix
transcription factor, putative"
/protein_id="AAL91233.1"
/db_xref="GI:19698995"
/translation="MDLSDIRNNNDTAATGAGAGQALVDASLSIVRSTPDPSTL
ATTSTATTATTKSTKDRHTKVDGGRIRMPALCAARVQLTRBLGHKSGDGTIEW
LLOQABPAIVATGTGTIPANFSTLSVLSRSGSTLSAPSKSVPLYGALGLTHQYD
EOGGGVFAHTSPLLFHQHLOHONONQDPVETIPEGNSFKRYSVDLSKEN
DDRQNKSLKSETSGPTAAWAPVPSRSGAGNTFMFLPVTAGNQKSSNN
NTAAGRAPMWPFPVNSAGGAGGGAATHFNAAGTGFSPMDQYRGSPLQGSFLAQ
PQPTQNLGLSMPDSNLGMLAALNSAYSRCGNANAEQANNAVEHQEKQSQSHDDDS
RENSNSSE"
399 a 310 c 388 g 382 t

gene

CDs

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

BASE COUNT 399 a 310 c 388 g 382 t

ORIGIN

Query Match 2.1%; Score 23; DB 8; Length 1479;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAGACCGTCACAGAA 242
Db 217 TCGACTAAGACCGTCACAGAA 239

FEATURES

source

1..85961
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/map="m1259"
/chromosome="1"
/clones="T18124"
1..1104
/gene="T18124.1"
/note="similar to unknown protein GI:2262162 from
(Arabidopsis thaliana)"
<1..>1104
/gene="T18124.1"
<1..>1104
/gene="T18124.1"
/codon_start=1
/product="hypothetical protein, 5' partial"
/protein_id="AAS0757.1"
/db_xref="GI:12321379"
/translation="KLFLDEELFVPRPSILVRSVCREWNTLFKRRFINKFNA
CGRPEIMLNTSHIVSIVDLKDNETIKYDRLDFHLSGCRVHLYGICDGNFMYSP
LNGGGVWNPFLFWRQTKAKAENTCGAIGYDGRPEKSKYKIGRSCSQMGKVTYD
TYSVFEATNAWYDTHETHEXPELMDSDGSRUSLGNLYXTAYASPHGTQYFLAMD
FSKLEKSRKTCVLPCKGEKSTHTRILSYAGDRFSLVLEQSKRTREIEIWTQDQI
GNGDDGDVVWIKFMTVSRPDPFILLISYISTSYFVNDIHGKSFVLCPCSKPKAAW
YIVRGDLCKIKIDQVLCFQSSVYPSLTIP"
complement(2159..3364)
/note="similar to auxin induced basic helix loop helix
transcription factor GI:5731257 from (Gossypium hirsutum)"
complement(<2159..>3364)
/gene="T18124.2"
complement(2159..3364)
/gene="T18124.2"
/codon_start=1
/product="auxin-induced basic helix-loop-helix
transcription factor putative"
/protein_id="AAS0759.1"
/db_xref="GI:12321381"
/translation="MDLSDIRNNNDTAATGAGAGQALVDASLSIVRSTPDPSTL
ATTSTATTATTKSTKDRHTKVDGGRIRMPALCAARVQLTRBLGHKSGDGTIEW
LLOQABPAIVATGTGTIPANFSTLSVLSRSGSTLSAPSKSVPLYGALGLTHQYD
EOGGGVFAHTSPLLFHQHLOHONONQDPVETIPEGNSFKRYSVDLSKEN
DDRQNKSLKSETSGPTAAWAPVPSRSGAGNTFMFLPVTAGNQKSSNN
NTAAGRAPMWPFPVNSAGGAGGGAATHFNAAGTGFSPMDQYRGSPLQGSFLAQ
PQPTQNLGLSMPDSNLGMLAALNSAYSRCGNANAEQANNAVEHQEKQSQSHDDDS
RENSNSSE"
complement(3567..3595)
/rpt_family="AT rich"
complement(4594..4638)

gene

mRNA

CDS

gene

mRNA

CDS

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

repet_region

```
gene
/rpt family="AT rich"
complement(5299..5853)
/gene="T18124.3"
/notes="contains similarity to bZIP transcriptional
activator RSG GI:8777512 from (Nicotiana tabacum)"
complement(join(5299..5357,5415..5651,5738..5864,
5968..6043,6135..6853))
/gene="T18124.3"
complement(join(5299..5357,5415..5651,5738..5864,
5968..6043,6135..6853))
/gene="T18124.3"
/codon start=1
/product="hypothetical protein"
/protein_id="AAG50761.1"
/db_xref="GI:12321383"
/translation="MASSSSGVNLMVPKHALPKIPPPSVASYSYIPGLI
GRHGKLNKTHQRTSSSHLVEELPFWDDLLNQPSRPAKCHRSSSDSYA
YLDVATNLSITLQNFYNTVSTQVQLDRLNKAQDAAPYSGASFQKRSQ
RSLVATACPSLWLPFAENGKGLGALYQSDATVTSERKNYAEFFSHDPKMLS
EENNPSPTVYADNTKAKQQAQRGRVRKQVISELERNVOTLQAEKVAEALD
FLNQRLILSMENKALKKLEISIAOEKLIKOYSTVLKIVILMSSVYVPEQVLEKEI
GLRALYQQOQOTKPSASGRATSKDLDSQSSLSLTKDSNCRDVSVMGPDSTI
RSDVLPPEHREF"
complement(8504..8535)
/rpt family="AT rich"
complement(8599..8627)
/rpt family="AT rich"
10149..11703
/gene="T18124.4"
/notes="Predicted by Genemark hm"
join(10149..10243,10536..11703)
/gene="T18124.4"
join(10149..10243,10536..11703)
/gene="T18124.4"
/codon start=1
/product="hypothetical protein"
/protein_id="AAG50763.1"
/db_xref="GI:12321385"
/translation="MEIQMRQMLMSLNLSKSRVSLTGGNHTAASADTFLKASRE
VLSFMRTLLALLFTSLWLLSKYGTATAPSKSVESDLPPELLPILLDLKQGLF
KNGKALFLSGDDDEVTSSYSQVETDMLVSASQDQSMVDPSTDFPAHSPH
IDSAEFIDRTLVGSIFTVQLNLODLPENLLKPNVYVVKSEYVNTWTKYETE
QKQSLVAGTKLIGITEDEAKALKLEDVLLPEPAAKSRSTYFKRTYRFDLNG
DNLLSVSRVFDVNGKSSGSEWEPENIPTRNKFENKIETVNDMSLESEKQ
KSLWLEKLEESVYNKAAEVEENMRMSKSKMVDLFLCKPLKGLRGRKQMS
KSGRAYWECALYGLRDEGVAVHQWNG"
complement(10335..10360)
/rpt family="GAGAA)n"
complement(12815..13163)
/gene="T18124.17"
/notes="pseudogene, putative heat shock transcription
factor HSF30"
/pseudo
complement(13870..13965)
/rpt family="AT rich"
complement(14034..14074)
/rpt family="AT rich"
complement(14182..14205)
/rpt family="AT rich"
complement(14901..14863)
/gene="T18124.5"
/notes="similar to copia type polyprotein GI:6850900 from
(Arabidopsis thaliana)"
complement(14901..14863)
/gene="T18124.5"
complement(14901..14863)
/gene="T18124.5"
/codon start=1
/product="copia-type polyprotein, putative"
/protein_id="AAG50765.1"
/db_xref="GI:12321387"
/translation="MAGNVPFPQVPLTKSNYNWLSRMAKILGAHDVMEIKGPIE
PENEGSLSQTKOGLDRSRDKKALCLLYQGLDEDTFEKVEATSAKEAWELRTSY
```

```
KGADQVKVRLQTLRGEFALOMKEGELYSYDSYFVSRVLTNNLRNGEKLDDVRIEMK
VLRSLDLKFEHIVTVEETKDLAMTIBOLLGSLQAYEEKKKKBDIVQVLNMOITK
EENGQYQRRGGGVRGGRGYNGRWPHEDNTNORGENSRGRGKGKPSRYDK
SSVYCNCGKFGHVAECKAPSKKFEKANYVEEKIQEDMLLMASYKQDQENHK
WYLDGSAHMCGRKKNFAELDSVRGNVALGDSKQEVKGNLILKNGDHQFIS
NYTIIISMKTNLSLQGLLEKGYDIRLKNNSIRDSNLTITKVPMSKRMFVLNIR
NDIAQLKMCYKEESSWHLRFGHNLGGLLESLRSEMVRGLFCINHPNOVCEGLLG
KQFMSPFKKSAQKPLELHTDVGPIKPSLGSNYFLFIDDFRKTWVYFLK
EKSEFEIKFKAHVEKESGLVITKMRSDRGEFTSKFELVCBNGTRROLTVPRPS
POONGVAERNRTILMARSMLKSLPKELNAAEACAVILNLSRPTKSVSGKTPQE
ANSRGKGVSHLVEGSIHAHVDPDEKSKLDDKSEKVIPIGYDNNKSKYLYNEDTK
KTIISNIPDEGEWDNSNEDNFPFPHFEDKFEFPEEPFPEEPFPEEPFPEEPF
QIEKECPMDQAEIKTWRNADDEIKSIQNDTWELTSLPNHGKAIVKWKYAK
KSGSVERKARLVAGYSQAGIDYDEVFAPVARLETVRLIISLAQKWKIHMND
VKSFLNGLEEVYIEQPGYIVKGEEDVLRLKALYGLQAAPRANTRIDKFKPE
KDFIKCPYHALYIKQEDILIACLYVDLLFTGNPMSMFEKFKETKSEMDIG
LMSYLGIEVQKQDNGIFITQRYAKVELKFKDDSNFVCTPMRGLSKSEBEG
VDPFTKSLVGLRYLTCRPLLVAGVGVVSMHPTTHKAKRLILYIKGVNF
GLHVSSTYDKLVGYSDDPGDGDVDDKSTSGSFVYIGDTATWMSKKQPIVLTSTCE
AEVVAATSCVCHAIWLRNLKELSPQEEPTKIFVDNKSALALAKNPVPHDSKHIDT
RHYIRECVSKXQVQLEYVKTHTQVADIFTPLKREDFTIOMESLILGVAKSLRGVGS
"
```

```
Query Match 2.1%; Score 23; DB 8; Length 85961;
Best Local Similarity 100.0%; Pred No. 4.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 220 TCAGCTAAAGACCGTCACACGAA 242
Db 3190 TCAGCTAAAGACCGTCACACGAA 3168
```

```
RESULT 10
AC079604/c
LOCUS
DEFINITION
Arabidopsis thaliana chromosome 1 BAC T15M6 genomic sequence,
complete sequence.
AC079604
AC079604.5 GI:12321249
VERSION
KEYWORDS
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
1 (bases 1 to 103637)
Lin.X., Kaul.S., Town.C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ranning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC T15M6 genomic sequence
Unpublished
2 (bases 1 to 103637)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (04-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
3 (bases 1 to 103637)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280852.
Address all correspondence to:at@tigr.org
BAC clone T15M6 is from Arabidopsis thaliana chromosome 1
clone.
The orientation of the sequence is from SP6 to T7 end of the BAC
clones.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of GlimmerM, see Mihaila Perlea
http://www.tigr.org/softlab/glimmerm_hcm/glimmerm.html, and
```


Genesplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@igr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

Source

1. 103637

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="1"

/map="2"

/clone="T15M6"

complement(1..663)

/gene="T15M6.26"

/note="pseudogene, hypothetical protein"

/pseudo

complement(1003..1088)

/rpt_family="AT_rich"

complement(1091..2595)

/gene="T15M6.25"

/note="pseudogene, hypothetical protein"

/pseudo

complement(3359..4729)

/gene="T15M6.24"

complement(<3359..>4729)

/gene="T15M6.24"

complement(3557..4729)

/gene="T15M6.24"

/codon_start=1

/product="unknown protein"

/protein_id="AAG50699.1"

/db_xref="GI:12321255"

/translation="MNTASAKTTRAGSNPEPDQLSPRSLKQKKWISVSLC
IFVLGDSLVMLLFFVQDNRSDODLQVGTWQALVQNAAPFLILFLFFP
SPQONQETTRFLSLILLYSLGLVAHSLKALGKLYANFGVLLIIGATLIF
TAIFAIIINRFRTRWIIISIGSILYVFGSPGDENEEFYSIQAWLTFPAASV
AFALSCLFQCEKVLKTKRYKKNVFRWVEMOICVSVFATVCLGLFASGENK
ELQGDHRPKGTIVYLSLIGLALSNQVWAGLGLVLYVSGVDFVHCTSLIVA
LPVILAFDNDPFRIGTLTATVVALGSIFFYTLKKNKKWVELYQTENNIDV"

complement(6149..6169)

/rpt_family="AT_rich"

complement(6713..6740)

/rpt_family="AT_rich"

complement(7047..7147)

/rpt_family="AT_rich"

complement(8659..8680)

/rpt_family="AT_rich"

complement(10552..10573)

/rpt_family="AT_rich"

complement(10742..10991)

/gene="T15M6.23"

/note="pseudogene, putative ribosomal protein"

/pseudo

complement(11760..11838)

/rpt_family="AT_rich"

complement(12070..17708)

/gene="T15M6.4"

/note="pseudogene, putative non-LTR retroelement reverse

transcriptase"

/pseudo

complement(18837..23614)

/gene="T15M6.22"

mRNA

/note="contains similarity to cell proliferation related
protein GI:9931486 from (Mus musculus)"
complement(1031..18837..19207,19528..19647,19786..19886,
20080..20291,20407..20787,20902..20989,21228..21288,
21891..22125,22758..22835,22952..23614))
/gene="T15M6.22"

CDS

complement(1031..18837..19207,19528..19647,19786..19886,
20080..20291,20407..20787,20902..20989,21228..21288,
21891..22125,22758..22835,22952..23614))
/gene="T15M6.22"

/codon_start=1

/product="hypothetical protein"

/protein_id="AAG50696.1"

/db_xref="GI:12321252"

/translation="MKRGRGHKKGKSKTINEQGNLNESTENAEQTSSESSAPVCE
SSQALEKMEVDAPSPAGSDVLDADNTAAKSVARVKVLTAKAPENETLRDDIDKS
ENSOALEKVPVVEKKEEFVRLPRKPVFLNVYTKGIRIKSSKAVDSSSVTBS
ADTVKVQDVVVGQDQTKSENSQASKAEIATISLOKEKKTQNTQYKQLELED
SLIVKIKMKEADPNVPUNPEALGIPDYFDIILKTPMDPGTLCNNFEKKNMSE
DYKDVYIWNCSKYNKKGDIYVDMKRVKQPMKIWTSAGLITQSAENTEDGGK
ASTQKSHKRGRRHKHDCMCAICVLKRRKRERDSGAQEESESPAGSPSVNNSVNM
GEDMDIDVKKPEQKTEIIVELSPVSKTORVIEKOEVEEENVESEENKTKANVE
DKTSIDRSMETGDEPNGAASEKLVVLALEGPKSTQNEEKEKRLQKQKRLSLE
RKEWRMKQKRFQVRNQLLSCELTLPDNNHNSVNGPHSLFRGGGNSRSLHK
AVESLMNLSLVALGASLFAMLDESGVFTLRAATSNASIFILSKAIGEYASL
ISKAVELTATDMLLKNCKSEKERRCDDAEVAGERSRLDYPFGDWGSELLR
RAVSIPVSPRTKPTESQSRQSLGLGVNELWSPVAETIRKSRSPQCLVQSGDGA
RQCGEIVLAMHCLCEVLLIRAPICYL"

19473..19512

/rpt_family="(A)n"

complement(24258..24296)

/rpt_family="AT_rich"

complement(25863..36393)

/gene="T15M6.5"

/note="pseudogene, putative cationic amino acid

transporter"

/pseudo

complement(30166..30215)

/rpt_family="(A)n"

complement(32458..32485)

/rpt_family="AT_rich"

complement(32989..33008)

/rpt_family="(TA)n"

complement(33253..33276)

/rpt_family="(TA)n"

complement(33753..33773)

/rpt_family="AT_rich"

complement(33893..33928)

/rpt_family="AT_rich"

37880..38594

/gene="T15M6.20"

/note="pseudogene, putative retroelement pol. polyprotein"

/pseudo

complement(39320..39376)

/rpt_family="AT_rich"

complement(39495..39523)

/rpt_family="AT_rich"

39684..39709

/rpt_family="(TAA)n"

complement(40447..49696)

/gene="T15M6.6"

/note="contains similarity to maleless protein (mle)

GI:157905 from (Drosophila melanogaster)"
complement(1031..18837..19207,19528..19647,19786..19886,
20080..20291,20407..20787,20902..20989,21228..21288,
21891..22125,22758..22835,22952..23614))
/gene="T15M6.22"

CDS

complement(join(40447..40527,40617..40739,40922..41017,
41217..41405,41490..41576,41787..41853,41997..42067,
42272..42409,42496..42597,42686..42769,42978..43034,
43117..43206,43297..43437,43824..43928,44069..44221,
44305..44408,44502..44607,44691..44874,44972..45057,
45292..45357,45493..45550,45745..45845,46042..46119,
46206..46328,46511..46587,46694..46882,46982..47093,
47463..47681,47870..48243,48329..48479,48788..48823,
49327..49132,49211..49316,49403..49696))
/gene="T15M6.6"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAGS0700.1"
/db_xref="GI:12321256"
/translation="MAPTKKPKQKQKNEIASSLPNSGHKKFKAPKLLISPENED

Query Match

Best Local Similarity 2.1%; Score 23; DB 8; Length 103637;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

220 TCAGCTAAGACCGTCACACGAA 242

|||||
|||||

DB

75568 TCAGCTAAGACCGTCACACGAA 75546

RESULT 11

LOCUS

AF352244 2134 bp mRNA linear INV 25-APR-2001

DEFINITION

Dermatophagoides farinae paramyosin-like allergen mRNA, partial

ACCESSION

AF352244

KEYWORDS

AF352244.1 GI:13785806

SOURCE

Dermatophagoides farinae (American house dust mite)

ORGANISM

Dermatophagoides farinae

REFERENCE

1 (bases 1 to 2134)

AUTHORS

Tsai, L.-C., Chao, P.-L., Shen, H.-D., Tang, R.-B., Chang, T.-C., Chang, Z.-N.,

TITLE

Isolation and characterization of a novel 98-kd Dermatophagoides

JOURNAL

J. Allergy Clin. Immunol. 102 (2), 295-303 (1998)

MEDLINE

9839068

PUBMED

9723675

REFERENCE

2 (bases 1 to 2134)

AUTHORS

Tsai, L., Sun, Y., Chao, P., Ng, H., Hung, M., Hsieh, K., Liaw, S. and

TITLE

Sequence analysis and expression of a cDNA clone encoding a 98-kDa

JOURNAL

allergen in Dermatophagoides farinae

PUBMED

Clin. Exp. Allergy 29 (12), 1606-1613 (1999)

AUTHORS

Tsai, L.-C., Chao, P.-L., Ng, H.-P. and Chua, K.-Y.

TITLE

Direct Submission

JOURNAL

Submitted (22-FEB-2001) Medical Research, Veterans General

Hospita-Taipei, No. 201, Sec. 2, Shih-Pai Road, Taipei, Taiwan

112, Republic of China

Location/Qualifiers

1. 2134

/organism="Dermatophagoides farinae"

/mol_type="mRNA"

/db_xref="taxon:6954"

/clone="Derf11"

57..>2134

/codon_start=1

/product="paramyosin-like allergen"

/protein_id="AAK3511.1"

/db_xref="GI:13785807"

/translation="NNKKRDSBELAKRLKLEDVHIESEETAHLRQKHAQIEMQDQ

LDLQKAKNKSDEKQKQFAEVFELLAQETANKKLTALKNKVEIKLEYTVHNLNIE

EINRTVIELTSHKQRLSQENTELIKEVHVKLQLDNANHLKTOIAQOLEDTRHRLBEE
ERKASLENHAHTLEVELESLKVLQDESEARLELERQLTKANGDAASWKSRYEABIQ
AHADVELERKWAQKISEYEQLEALLNKCSLEKQKSLQSEVFLINDLEKATRH
AQLAEKRVADLEKINDLKNKLREVTMLMEQAKOELRVKIAELQLOHVEKLRDORD
QILARENKLLDDEAKSOLANDARRHIEQEIKELENERDELSSAAYKEATLRQOE
EAKNQRLIAQLAVRHDIYERKLAQKDEFEIALRKQYQIELEQNLMLAEAAKLTVEI
ARLKCKYQOQITELSELSDAANKANIDLQKTIKKQALQITELQAHYDEVHROQQOAVD
CLGVTQRCAQLQAELEEMRLALEQANRAKQAEQLEHFAVRVNVELTTINVLAKAK
SKLESEFSAQADYDEVHKLRLISDRVQKLTIELKSTKDLITTEQERIVKLETVKKS
LEQEVRTLHVRIEVEANALAGGKRVIAKLESIRIDVETEVEERRRHAETDQLRKK
DHRVKELLQ"

BASE COUNT 838 a 354 c 462 g 480 t

ORIGIN

Query Match 2.1%; Score 22; DB 3; Length 2134;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

949 AGTTCATCGTCATTGCAACAA 970

DB

1550 AGTTCATCGTCATTGCAACAA 1571

RESULT 12

LOCUS

AC124912

DEFINITION

Papio anubis clone RP41-205M12, WORKING DRAFT SEQUENCE, 4 ordered

PIECES.

AC124912

ACCESSION

AC124912.2 GI:21844567

VERSION

HTG; HTGS PHASE2; HTGS DRAFT.

KEYWORDS

Papio anubis (olive baboon)

SOURCE

ORGANISM

REFERENCE

AUTHORS

Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S. M.,

Benjamin, B., Blakesley, R. W., Bouffard, G. G., Breen, K., Brinkley, C.,

Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, J.,

Haghighi, P., Hansen, N., Ho, S.-L., Idol, J. R., Karlins, E., Laric, P.,

Lee-lin, S.-O., Legaspi, R., Maduro, Q. L., Maduro, V. B.,

Margulies, E. H., Masello, C., Maskeri, B., Mastrian, S. D.,

McCloskey, J. C., McDowell, J., Paguirigan, C., Pearson, R.,

Portnoy, M. E., Prasad, A., Schueler, M. G., Stantripop, S., Thomas, J. W.,

Thomas, P. J., Touchman, J. W., Teursson, C., Vogt, J. L., Walker, M. A.,

Wetherby, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D.

NISC Comparative Sequencing Initiative

Unpublished

REFERENCE

2 (bases 1 to 149452)

AUTHORS

Green, E. D.

TITLE

Direct Submission

JOURNAL

Submitted (19-JUN-2002) NIH Intramural Sequencing Center, 8717

Grovermont Circle, Gaithersburg, MD 20877, USA

REFERENCE

3 (bases 1 to 149452)

AUTHORS

Green, E. D.

TITLE

Direct Submission

JOURNAL

Submitted (16-JUL-2002) NIH Intramural Sequencing Center, 8717

Grovermont Circle, Gaithersburg, MD 20877, USA

COMMENT

On Jul 16, 2002 this sequence version replaced gi:21455384.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: http://www.nisc.nih.gov

Contact: nisc.zoo@nhgri.nih.gov

----- Project Information

Center project name: cyu

Center clone name: 205M12

CDS

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair

data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 148556 bases at least Q40
Consensus quality: 148787 bases at least Q30
Consensus quality: 148973 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 149152; sum-of-contigs
Quality coverage: 11.35x in Q20 bases; agarose-fp
Quality coverage: 11.42x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved

1 18748 18847: contig of 18747 bp in length
1 18848 18848: gap of unknown length
1 127485 127584: gap of 108637 bp in length
1 127585 147017: contig of unknown length
1 147018 147117: gap of 19433 bp in length
1 147118 149452: contig of 2335 bp in length.

FEATURES

Source

1. 149452
/organism="Papio anubis"
/mol_type="genomic DNA"
/db_xref="taxon:9555"
/clones="RP41-205M12"
/clone_lib="RP41"

misc_feature

1. 44256
/note="Clone overlaps with GenBank Accession Number AC125510 clone RP41-382N23 (center project name cyt)"

misc_feature

1. 18747
/note="assembly_fragment"
clone end:SP6
vector side:left

misc_feature

18848 - 127484
/note="assembly_fragment"

misc_feature

127585 - 147017
/note="assembly_fragment"

misc_feature

147118 - 149452
/note="assembly_fragment"
clone end:T7
vector side:right

BASE COUNT

46580 a 31101 c 29868 g 41603 t 300 others

Query Match

Best Local Similarity 2.1%; Score 22; DB 2; Length 149452;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

473 TGGGTGAATCTGATCAAGAA 494

Db

35684 TGGGTGAATCTGATCAAGAA 35705

RESULT 13

AC125510

LOCUS

AC125510 170737 bp DNA linear HTG 20-JUL-2002

DEFINITION

Papio anubis clone RP41-382N23, WORKING DRAFT SEQUENCE, 4 ordered pieces.

ACCESSION

AC125510

VERSION

HTG; HTGS PHASE2; HTGS DRAFT.

KEYWORDS

Papio anubis (olive baboon)

SOURCE

Papio anubis

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Papio.

REFERENCE

1 (bases 1 to 170737)

AUTHORS

Akhter N., Antonellis A., Ayele K., Beckstrom-Sternberg S.M., Benjamin B., Blakesley R.W., Bouffard G.G., Breen K., Brinkley C., Brooks S., Dietrich N.L., Granite S., Guan X., Gupta J., Haghighi P., Hansen N., Ho S.-L., Idol J.R., Karlins E., Laric P., Lee-Lin S.-Q., Legaspi R., Maduro Q.L., Maduro V.B., Marquies E.H., Masello C., Maskeri B., Mastroian S.D., McCloskey J.C., McDowell J., Paguirigan C., Pearson R., Portnoy M.E., Prasad A., Schueler M.G., Stantropop S., Thomas J.W., Thomas P.J., Touchman J.W., Tsurgon C., Vogt J.L., Walker M.A., Wetherby K.D., Wiggins L., Young A., Zhang L.-H. and Green E.D.

TITLE

Submitted (28-JUN-2002) NIH Intramural Sequencing Center, 8717

JOURNAL

Green E.D.

REFERENCE

2 (bases 1 to 170737)

AUTHORS

Green E.D.

JOURNAL

Submitted (28-JUN-2002) NIH Intramural Sequencing Center, 8717

TITLE

Submitted (20-JUL-2002) NIH Intramural Sequencing Center, 8717

JOURNAL

Submitted (20-JUL-2002) NIH Intramural Sequencing Center, 8717

COMMENT

On Jul 20, 2002 this sequence version replaced gi:21622721.

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: http://www.nisc.nih.gov

Contact: nisc.zoo@nih.gov

Center project name: cyt

Center clone name: 382N23

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 169910 bases at least Q40
Consensus quality: 170146 bases at least Q30
Consensus quality: 170296 bases at least Q20
Insert size: 162000; agarose-fp
Insert size: 170437; sum-of-contigs
Quality coverage: 11.73x in Q20 bases; agarose-fp
Quality coverage: 11.15x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 *
 * 89919: contig of 89919 bp in length
 * 90020: gap of unknown length
 * 146139: contig of 56119 bp in length
 * 146139: gap of unknown length
 * 146239: contig of 17872 bp in length
 * 164111: gap of unknown length
 * 164211: contig of 6527 bp in length.

FEATURES

source
 1. 170737
 /organism="Papio anubis"
 /mol_type="genomic DNA"
 /db_xref="taxon:9555"
 /clone="RP41-382N23"
 /clone_lib="RP41"
 misc_feature
 1. 89919
 /notes="assembly_fragment
 clone end:SP6
 vector side:left"
 misc_feature
 1. 38014
 /notes="clone overlaps with GenBank Accession Number
 AC12511 clone RP41-427L1 (center project name cys)"
 90020. 146138
 /notes="assembly_fragment"
 132730. 170737
 /note="clone overlaps with GenBank Accession Number
 AC124912 clone RP41-205M12 (center project name cyu)"
 146239. 164110
 /note="assembly_fragment"
 164211. 170737
 /note="assembly_fragment
 clone end:T7
 vector side:right"
 ASE COUNT 55152 a 35045 c 32826 g 47413 t 300 others
 RIGIN

Query Match 2.1%; Score 22; DB 2; Length 170737;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 473 TGGGTGAAATCTGATGAGAA 494
 b 162194 TGGGTGAAATCTGATGAGAA 162215

RESULT 14
 C051640
 OCUS AC051640 158514 bp DNA linear HTG 24-AUG-2002
 SEQUENCE, 25 unordered pieces.
 AC051640
 AC051640.4 GI:9966320
 HTG; HTGS PHASE1; HTGS_DRAFT.
 BYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 158514)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 9, clone RP11-161B4
 Unpublished
 2 (bases 1 to 158514)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hags,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 158514)

REFERENCE

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
 Bouckgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
 FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
 Graham,L., Grand-Pierre,N., Hags,B., Heaford,A., Horton,L.,
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
 Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
 O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Soudnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zimmer,A. and Zody,M.

TITLE

JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 4, 2000 this sequence version replaced gi:7960427.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: NIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8703

Center clone name: 161_B_4

----- Summary Statistics

Sequencing vector: M13; M77815; 93% of reads

Sequencing vector: Plasmid; n/a; %-0.1% of reads

6.50344299923489Chemistry: Dye-terminator Big Dye; 100% of

reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136255 bases at least Q40

Consensus quality: 147839 bases at least Q30

Consensus quality: 152536 bases at least Q20

Insert size: 159000; agarose-fp

Insert size: 156114; sum-of-contigs

Quality coverage: 3.4 in Q20 bases; agarose-fp

Quality cov.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1783: contig of 1783 bp in length


```
* 1784 1883: gap of 100 bp
* 1884 3231: contig of 1348 bp in length
* 3232 3331: gap of 100 bp
* 3332 4352: contig of 1021 bp in length
* 4353 4452: gap of 100 bp
* 4453 5702: contig of 1250 bp in length
* 5703 5803: gap of 100 bp
* 5804 8393: contig of 2591 bp in length
* 8394 8493: gap of 100 bp
* 8494 10272: contig of 1779 bp in length
* 10273 10372: gap of 100 bp
* 10373 12752: contig of 2380 bp in length
* 12753 12852: gap of 100 bp
* 12853 15516: contig of 2664 bp in length
* 15517 15616: gap of 100 bp
* 15617 15622: contig of 5946 bp in length
* 15623 19663: gap of 100 bp
* 19664 33483: contig of 13721 bp in length
* 33484 38548: contig of 5065 bp in length
* 38549 38648: gap of 100 bp
* 38649 43137: contig of 4489 bp in length
* 43138 43237: gap of 100 bp
* 43238 48586: contig of 5349 bp in length
* 48587 48686: gap of 100 bp
* 48687 55083: contig of 6397 bp in length
* 55084 55183: gap of 100 bp
* 55184 61762: contig of 6579 bp in length
* 61763 61862: gap of 100 bp
* 61863 72589: contig of 10727 bp in length
* 72590 72689: gap of 100 bp
* 72690 80806: contig of 8117 bp in length
* 80807 80906: gap of 100 bp
* 80907 88472: contig of 7566 bp in length
* 88473 88572: gap of 100 bp
* 88573 88622: contig of 10052 bp in length
* 88623 98724: gap of 100 bp
* 98725 107382: contig of 8658 bp in length
* 107383 107482: gap of 100 bp
* 107483 117644: contig of 10162 bp in length
* 117645 117744: gap of 100 bp
* 117745 128355: contig of 10611 bp in length
* 128356 128456: gap of 100 bp
* 128457 140324: contig of 11869 bp in length
* 140325 140424: gap of 100 bp
* 140425 153983: contig of 13559 bp in length
* 153984 154083: gap of 100 bp
* 154084 158514: contig of 4431 bp in length.
FEATURES
    Location/Qualifiers
        1..158514
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="9"
            /map="9"
            /clone="RP11-161B4"
            /clone_lib="RP11-11 Human Male BAC"
        1..1783
            /note="assembly_fragment
            clone end:SP6
            vector side:left"
        1884..3231
            /note="assembly_fragment"
        3332..4352
            /note="assembly_fragment"
        4453..5702
            /note="assembly_fragment"
        5803..8393
            /note="assembly_fragment"
        8494..10272
            /note="assembly_fragment"
        10373..12752
            /note="assembly_fragment"
```

```
misc_feature 12853..15516
              /note="assembly_fragment"
misc_feature 15617..19562
              /note="assembly_fragment"
misc_feature 19663..33383
              /note="assembly_fragment"
misc_feature 33484..38548
              /note="assembly_fragment"
misc_feature 38649..43137
              /note="assembly_fragment"
misc_feature 43238..48586
              /note="assembly_fragment"
misc_feature 48687..55083
              /note="assembly_fragment"
misc_feature 55184..61762
              /note="assembly_fragment"
misc_feature 61863..72589
              /note="assembly_fragment"
misc_feature 72690..80806
              /note="assembly_fragment"
misc_feature 80907..88472
              /note="assembly_fragment"
misc_feature 88573..98624
              /note="assembly_fragment"
misc_feature 98725..107382
              /note="assembly_fragment"
misc_feature 107483..117644
              /note="assembly_fragment"
misc_feature 117745..128355
              /note="assembly_fragment"

Query Match      2.0%; Score 21; DB 2; Length 158514;
Best Local Similarity 100.0%; Pred. NO. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      244 GTTGAAGGAGAGGAGGAGGAGG 364
      |||
Db      149299 GTTGAAGGAGAGGAGGAGGAGG 149319

RESULT 15
AL499604
LOCUS
DEFINITION Human DNA sequence from clone RP11-23B15 on chromosome 9, complete
            sequence.
ACCESSION AL499604 AC015928
VERSION AL499604.9 GI:13274855
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 160796)
AUTHORS Garner, P.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
COMMENT On or before May 15, 2001 this sequence version replaced
            GI:7382602, GI:13273876.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e. phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
```

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; SWI, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr-9> RP11-23B15 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-23B15 The true right end of clone RP11-54606 is at 35092 in this sequence.

FEATURES

Source

Location/Qualifiers
 1..160796
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-23B15"
 /clone_lib="RP11-11.1"

repeat_region
 289..503
 /note="MER20 repeat: matches 1..214 of consensus"
 629..915
 /note="AluJo repeat: matches 1..286 of consensus"
 2346..2870
 /note="MLT1G repeat: matches 14..536 of consensus"
 2876..3130
 /note="LIMC5 repeat: matches 6976..7233 of consensus"
 3182..3503
 /note="LIMC5 repeat: matches 7282..7595 of consensus"
 3504..3816
 /note="AluX repeat: matches 1..310 of consensus"
 3819..4135
 /note="AluX repeat: matches 2..310 of consensus"
 4136..4540
 /note="LIMC5 repeat: matches 7595..7936 of consensus"
 5076..5127
 /note="MIR repeat: matches 79..130 of consensus"
 5144..5373
 /note="MIR repeat: matches 2..260 of consensus"
 6766..6870
 /note="MER86 repeat: matches 25..132 of consensus"
 6895..7062
 /note="LIPB3 repeat: matches 5964..6141 of consensus"
 7468..7858
 /note="L2 repeat: matches 2293..2709 of consensus"
 8086..8450
 /note="MER47A repeat: matches 2..366 of consensus"
 8607..8714
 /note="MER81 repeat: matches 1..107 of consensus"
 9115..9302
 /note="L2 repeat: matches 1610..1828 of consensus"
 9462..9713
 /note="L2 repeat: matches 2168..2419 of consensus"
 9714..10159
 /note="MLT2CB repeat: matches 1..501 of consensus"
 10160..12130
 /note="L2 repeat: matches 200..2168 of consensus"
 12131..12507
 /note="THEIC repeat: matches 1..371 of consensus"
 12508..12552
 /note="L2 repeat: matches 56..200 of consensus"
 13576..13692
 /note="MER63 repeat: matches 14..131 of consensus"
 13762..14168
 /note="MLT2FA repeat: matches 1..400 of consensus"
 14228..14920
 /note="SVA repeat: matches 552..1259 of consensus"
 14273..15366
 /note="CpG island"

/evidence=not experimental
 14921..15936
 /note="SVA repeat: matches 9..954 of consensus"
 15940..16092
 /note="LIM4 repeat: matches 3221..3374 of consensus"
 16351..16438
 /note="AluX repeat: matches 1..289 of consensus"
 16534..16691
 /note="MIR repeat: matches 18..184 of consensus"
 17093..17158
 /note="33 copies 2 mer aa 69% conserved"
 17377..17426
 /note="25 copies 2 mer aa 98% conserved"
 18077..18230
 /note="Charlie4 repeat: matches 1795..1952 of consensus"
 18294..18606
 /note="AluSq repeat: matches 1..313 of consensus"
 18610..18663
 /note="27 copies 2 mer ag 72% conserved"
 18666..18971
 /note="LIMB8 repeat: matches 5875..6168 of consensus"
 19056..19295
 /note="LIMB8 repeat: matches 1655..1888 of consensus"
 19285..19596
 /note="LIM4 repeat: matches 5433..5742 of consensus"
 19597..19715
 /note="MLT11 repeat: matches 19..137 of consensus"
 19772..20057
 /note="AluSp/q repeat: matches 212..297 of consensus"
 20319..20473
 /note="MIR repeat: matches 59..210 of consensus"
 20925..21126
 /note="LIMB2 repeat: matches 5953..6164 of consensus"
 21155..21464
 /note="AluSq repeat: matches 1..308 of consensus"
 21485..21679
 /note="5 copies 39 mer 69% conserved"
 21730..22258
 /note="CpG island"
 22197..22386
 /note="MIR repeat: matches 47..240 of consensus"
 22409..22522
 /note="MIR repeat: matches 25..145 of consensus"
 22741..22808
 /note="34 copies 2 mer ca 72% conserved"
 23245..23554
 /note="AluSq repeat: matches 1..308 of consensus"
 23663..23826
 /note="L2 repeat: matches 1441..1600 of consensus"
 24288..24416
 /note="MIR repeat: matches 22..145 of consensus"
 24467..24492
 /note="13 copies 2 mer ac 96% conserved"
 24712..24763
 /note="36 copies 2 mer tt 76% conserved"
 24765..24893
 /note="MIR repeat: matches 52..188 of consensus"
 25072..25317
 /note="MIR repeat: matches 1..250 of consensus"
 25948..25993
 /note="2 copies 23 mer 100% conserved"
 26307..26466
 /note="MLTID repeat: matches 5..166 of consensus"
 26543..26701
 /note="MIR repeat: matches 25..183 of consensus"
 26726..26839
 /note="MLT11 repeat: matches 19..134 of consensus"
 27352..27501
 /note="MIR repeat: matches 20..184 of consensus"
 27509..27576
 /note="34 copies 2 mer gg 66% conserved"
 27681..27925

```
repeat_region /note="L2 repeat: matches 2407. .2668 of consensus"
27938.28038
repeat_region /note="MIR repeat: matches 100. .262 of consensus"
29269.29580
repeat_region /note="AluSg repeat: matches 1. .313 of consensus"
29600.29946
repeat_region /note="L1MB8 repeat: matches 5817. .6171 of consensus"
29953.30009
repeat_region /note="L2 repeat: matches 2671. .2729 of consensus"
30708.31108
repeat_region /note="L1MB4 repeat: matches 5781. .6183 of consensus"
31109.31166
repeat_region /note="29 copies 2 mer at 75% conserved"
31131.31208
repeat_region /note="2 copies 39 mer 83% conserved"
31165.31290
repeat_region /note="3 copies 42 mer 79% conserved"
31172.31283
repeat_region /note="56 copies 2 mer at 84% conserved"
31294.32004
repeat_region /note="L1MB4 repeat: matches 5081. .5786 of consensus"
32005.32312
repeat_region /note="Aluub repeat: matches 1. .308 of consensus"

Query Match 2.0%; Score 21; DB 9; Length 160796;
Best Local Similarity 100.0%; Pred.No. 41;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 GTTGAAGGAGGAGGAGGAGG 264
Db 12691 GTTGAAGGAGGAGGAGGAGG 12711
```

Search completed: February 2, 2004, 07:55:25
Job time : 4077.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DN nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:59:06 ; Search time 313.739 Seconds
(without alignments)
9214.976 Million cell updates/sec

File: US-09-938-842A-1034

Perfect score: 1071
Sequence: 1 atgggacaattcagaagct.....caggtcatcgcaaccactga 1071

Scoring table: OLIGO NUC
Gapop_60.0, Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N Geneseq 19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length DB	ID	Description
1	1071	100.0	1071	24	Arabidopsis thalia
2	289	27.0	460	25	Arabidopsis thalia
3	286	26.7	453	24	Arabidopsis thalia
4	21	2.0	755	24	Oligonucleotide fo
5	21	2.0	755	24	Oligonucleotide fo
6	20	1.9	453	22	Human Maguin-2 hom
7	19	1.8	416	21	Zea mays DNA fragm
8	19	1.8	499	21	Arabidopsis thalia

c	9	19	1.8	600	24	ABQ43702	Oligonucleotide fo
c	10	19	1.8	600	24	ABQ43703	Oligonucleotide fo
c	11	19	1.8	908	24	ABL99906	Human secretory po
c	12	19	1.8	1156	21	ABC39595	Arabidopsis thalia
c	13	19	1.8	1893	21	AAC43180	Arabidopsis thalia
c	14	19	1.8	2794	24	ABL99921	Human secretory po
c	15	19	1.8	15987	23	ABU44198	Drosophila melanog
c	16	19	1.8	34094	20	AA30163	Complete nucleotid
c	17	18	1.7	277	24	ABL75342	Corn tassal-derive
c	18	18	1.7	424	22	AA325423	Human ovarian PCR-
c	19	18	1.7	432	25	ABX55556	Bovine EST associa
c	20	18	1.7	475	24	ABL93299	Arabidopsis thalia
c	21	18	1.7	605	23	ABL17167	Drosophila melanog
c	22	18	1.7	626	24	ABG66162	Arabidopsis thalia
c	23	18	1.7	645	21	AAC45826	Arabidopsis thalia
c	24	18	1.7	674	24	ABQ33616	Oligonucleotide fo
c	25	18	1.7	674	24	ABQ33617	Oligonucleotide fo
c	26	18	1.7	692	23	AA380632	DNA encoding novel
c	27	18	1.7	721	22	AA196937	Human neuroblastom
c	28	18	1.7	848	21	AAC45001	Arabidopsis thalia
c	29	18	1.7	852	24	ABQ44878	Oligonucleotide fo
c	30	18	1.7	852	24	ABQ44879	Oligonucleotide fo
c	31	18	1.7	954	21	AAC49097	Arabidopsis thalia
c	32	18	1.7	1012	21	AAC34099	Arabidopsis thalia
c	33	18	1.7	1029	24	ABQ45262	Oligonucleotide fo
c	34	18	1.7	1029	24	ABQ45263	Oligonucleotide fo
c	35	18	1.7	1049	21	ABC39576	Arabidopsis thalia
c	36	18	1.7	1120	21	AAC41903	Arabidopsis thalia
c	37	18	1.7	1390	22	AA60982	P. putida K12440-a
c	38	18	1.7	1399	24	ABQ47386	Oligonucleotide fo
c	39	18	1.7	1399	24	ABQ47387	Oligonucleotide fo
c	40	18	1.7	1449	24	ABQ13552	Oligonucleotide fo
c	41	18	1.7	1449	24	ABQ13553	Oligonucleotide fo
c	42	18	1.7	1452	21	ABC39557	Arabidopsis thalia
c	43	18	1.7	1733	22	AAK71547	Human immune/haema
c	44	18	1.7	1733	22	AAK71548	Human immune/haema
c	45	18	1.7	2271	23	ABL19369	Drosophila melanog

ALIGNMENTS

RESULT 1
ABZ13229
ID ABZ13229 standard; DNA; 1071 BP.
XX AC ABZ13229;
XX DT 21-JAN-2003 (first entry)
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1034.
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX OS Arabidopsis thaliana.
XX PN WO200216555-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26585.
XX PR 24-AUG-2000; 2000US-227866P.
XX PR 26-JAN-2001; 2001US-264647P.
XX PR 22-JUN-2001; 2001US-300111P.
XX (SCRI) SCRIPPS RES INST.
XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX PI Harper JF, Krepis J, Wang X, Zhu T;
XX DR WPI; 2002-304127/34.

PT Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
XX
PS Claim 144; SEQ ID NO 1034; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
CC in methods of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 1071 BP; 287 A; 288 C; 256 G; 240 T; 0 other;

Query Match 100.0%; Score 1071; DB 24; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCGCAATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCAAGCGGTT 60
DB 1 ATGGCGCAATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCAAGCGGTT 60
QY 61 GATCTAACCATCATCAACGGCGTCAGAAAGCTCGAACTTCAAGCAAT 120
DB 61 GATCTAACCATCATCAACGGCGTCAGAAAGCTCGAACTTCAAGCAAT 120
QY 121 CCCACGTGAGTCTCGAGCCCAAGCGCGCGTGCATGCGCTTCAATGCTTTA 180
DB 121 CCCACGTGAGTCTCGAGCCCAAGCGCGCGTGCATGCGCTTCAATGCTTTA 180
QY 181 GCTCCACCGTCTTCGACGAGCACCACCAATTAAGAGAGCTTCGACTAAGACCGTCA 240
DB 181 GCTCCACCGTCTTCGACGAGCACCACCAATTAAGAGAGCTTCGACTAAGACCGTCA 240
QY 241 AAGCTTCAAGGAGGAGAGATACGATGCTGCGCAAGTTCGGTGGAGATTTT 300
DB 241 AAGCTTCAAGGAGGAGAGATACGATGCTGCGCAAGTTCGGTGGAGATTTT 300
QY 301 CAATTAACCTGAGAGTTAGTCAAAATCCGAGCAAGTTCGGTGGAGATTTT 360
DB 301 CAATTAACCTGAGAGTTAGTCAAAATCCGAGCAAGTTCGGTGGAGATTTT 360
QY 361 AACGCTGAGCGCGGATTAAGCGCCACCGGTCACGAGTTCGGTGGAGATTTT 420
DB 361 AACGCTGAGCGCGGATTAAGCGCCACCGGTCACGAGTTCGGTGGAGATTTT 420
QY 421 TCGGTTAAGCGAAGTAAATCCGAGCAAGTTCGATGCTGATGCTGATG 480
DB 421 TCGGTTAAGCGAAGTAAATCCGAGCAAGTTCGATGCTGATGCTGATG 480
QY 481 AATCTGATGAAGAGAAACGTAACGACCTTCTAAGAGTATATAGACATAGCGAC 540
DB 481 AATCTGATGAAGAGAAACGTAACGACCTTCTAAGAGTATATAGACATAGCGAC 540
QY 541 GCGGTTTCACTTCCGCTTGTAGTCCAAATTCGAGCAAGTTCGATGCTGATG 600
DB 541 GCGGTTTCACTTCCGCTTGTAGTCCAAATTCGAGCAAGTTCGATGCTGATG 600
QY 601 CAAGCTCTGCATCATCACTGTGGTTCAGCAACTTCTCGCAAGGATGATCGATG 660
DB 601 CAAGCTCTGCATCATCACTGTGGTTCAGCAACTTCTCGCAAGGATGATCGATG 660
QY 661 TGGGCTATTCATCAACGCAATGATTCGACGGTGGAGCTTCTTCTTGTATCAAA 720

DB 661 TGGGCTATTCATCAACGCAATGATTCGACGGTGGAGCTTCTTCTTGTATCAAA 720
QY 721 ATCGTGTGCTCGGTCGAATCAGCTCAGTTATAGCTTTTCCGCGCGGCTTCCCG 780
DB 721 ATCGTGTGCTCGGTCGAATCAGCTCAGTTATAGCTTTTCCGCGCGGCTTCCCG 780
QY 781 TCGTCTTAGTTCGCGCTGTTCAACAGAGCTTCCAGAGTGGCTAGACCACTCTTTCAA 840
DB 781 TCGTCTTAGTTCGCGCTGTTCAACAGAGCTTCCAGAGTGGCTAGACCACTCTTTCAA 840
QY 841 GTTGTTCCTCAAGCAGCGGCTTTGATCCGTTTTCAGAGCTTAGCGGTTCAATTAAGA 900
DB 841 GTTGTTCCTCAAGCAGCGGCTTTGATCCGTTTTCAGAGCTTAGCGGTTCAATTAAGA 900
QY 901 GCGAGTCTGCTTATGCTTCCAGAGTCAAGCTCAAGCTTAACACCGGTAGTTCATCGTCA 960
DB 901 GCGAGTCTGCTTATGCTTCCAGAGTCAAGCTCAAGCTTAACACCGGTAGTTCATCGTCA 960
QY 961 ATTGCAACAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1020
DB 961 ATTGCAACAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1020
QY 1021 GAGCTTCAACAGTTCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1071
DB 1021 GAGCTTCAACAGTTCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1071
RESULT 2
ABX61156/c
ID ABX61156 standard; DNA; 460 BP.
XX
AC ABX61156;
XX
DT 26-FEB-2003 (first entry)
XX
DE Arabidopsis thaliana polynucleotide #502.
XX
KW Thale cress; gene; ds; genetic manipulation; plant; biosynthesis;
KW genetic modification; environmental stress; disease resistance;
KW fungicide; insecticide; stress tolerance.
XX
OS Arabidopsis thaliana.
XX
PN US2002142319-A1.
XX
PD 03-OCT-2002.
XX
PR 07-AUG-2001; 2001US-0924035.
XX
PR 13-AUG-1999; 99US-149784P.
XX
PR 11-AUG-2000; 2000US-0638258.
XX
PA (GORL//) GORLACH J.
PA (ANY//) AN Y.
PA (HAMI//) HAMILTON C M.
PA (PRIC//) PRICE J L.
PA (HARG//) HARGISS T R.
PA (YUY//) YU Y.
PA (RAME//) RAMEAKA J G.
PA (PAGE//) PAGE A.
PA (MATH//) MATHW A V.
PA (LEDF//) LEDFORD B L.
PA (WOES//) WOESSNER J P.
PA (HAAS//) HAAS W D.
PA (GARC//) GARCIA C A.
PI Gorlach J, An Y, Hamilton CM, Price JL, Hargiss TR, Yu Y,
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,
PI Garcia CA,
XX
DR WPI; 2003-102509/09.
XX
PT Novel Arabidopsis thaliana nucleic acid useful for constructing a

transgenic plant with enhanced disease resistance and enhanced traits of interest, as probes, and in diagnosis and screening purposes -

Claim 1; Page 158; 277pp; English.

The invention relates to Arabidopsis thaliana nucleic acid sequences. The DNA sequences and the polypeptides they encode are useful for identifying homologous or related genes, for producing compositions that modulate the expression or function of the polypeptides, for mapping associated regions of the protein, in diagnosis, for studying associated physiological pathways, for genetic manipulation of cells, preferably plant cells, in screening assays of various plant strains to determine the strains that are capable of withstanding a particular disease or environmental stress, for enhancing or inhibiting production of biosynthetic products in plants and to create genetically modified and transgenic organisms, such as plant cells and plants. Transgenic plants are useful for introducing or improving disease resistance and stress tolerance in plants, screening biologically active agents, such as fungicides and insecticides, and for identifying factors involved in biosynthetic pathways of nutritional, commercial or medicinal value. Sequences ABX0655-ABX61554 represent Arabidopsis thaliana polynucleotides of the invention.

Sequence 460 BP; 113 A; 103 C; 134 G; 109 T; 1 other;

Query Match 27.0%; Score 289; DB 25; Length 460;

Best Local Similarity 99.6%; Pred. No. 1.4e-131;

Matches 459; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

600 GCAAGCTTGGCATCATCTGCTGGCTCAGCAACTTTCGGCGAAGAAATGATCCGAT 659

460 GCAAGCTTGGCATCATCTGCTGGCTCAGCAACTTTCGGCGAAGAAATGATCCGAT 401

660 GTGGGCTATTCCATCAAGCAATCATTCGGAGCTCGAGCTTCTTGTATCCACA 719

400 GTGGGCTATTCCATCAAGCAATCATTCGGAGCTCGAGCTTCTTGTATCCACA 341

720 AATCGCTGGTCCGTCGAATCAAGCTCACTTATTAGCTTTTCCGCGCGCTGCTCGCC 779

340 AATCGCTGGTCCGTCGAATCAAGCTCACTTATTAGCTTTTCCGCGCGCTGCTCGCC 282

780 GTGCTCTTACGTCGCGCTGTTCAACAGCTTCCAGATGCTAGACACCTCTTTACA 839

281 GTGCTCTTACGTCGCGCTGTTCAACAGCTTCCAGATGCTAGACACCTCTTTACA 222

840 AGTTGTTCCAGCAGCGCTTGTATCGCTTTCAGACGTTAGCGTTCGAATTTCAAG 899

221 AGTTGTTCCAGCAGCGCTTGTATCGCTTTCAGACGTTAGCGTTCGAATTTCAAG 162

900 AGCGAGCTGGTTATGGTCCGAGCTCAAGCTCAGGCGTAACAACCGGTATTCATGTC 959

161 AGCGAGCTGGTTATGGTCCGAGCTCAAGCTCAGGCGTAACAACCGGTATTCATGTC 102

960 AATTCCAAACAACAGCAGCGCAGCTCAGAGATTTCTCCCTAGATATACGGAACA 1019

101 AATTCCAAACAACAGCAGCGCAGCTCAGAGATTTCTCCCTAGATATACGGAACA 42

1020 AGAGTTTCCAGTTTATGACCAACCAACAGCAGCTCAT 1060

41 AGAGTTTCCAGTTTATGACCAACCAACAGCAGCTCAT 1

BL93850/c

D ABU93850 standard; cDNA; 453 BP.

X ABU93850;

X 10-JUN-2002 (first entry)

X Arabidopsis thaliana nucleic acid sequence Ref:2027615 SEQ ID NO:615.

X Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;

genetic modification; gene; ss.

Arabidopsis thaliana.

US2002023280-A1.

21-FEB-2002.

26-JAN-2001; 2001US-0770444.

27-JAN-2000; 2000US-178502P.

(GORLACH J. CORLACH J.

(ANYI// AN Y.

(HAMI// HAMILTON C M.

(PRIC// PRICE J L.

(RAIN// RAINES T M.

(YUY// YU Y.

(RAME// RAMEAKA J G.

(PAGE// PAGE A.

(MATH// MATHW A V.

(LEDF// LEDFORD B L.

(WOES// WOESSNER J P.

(HAAS// HAAS W D.

(GARC// GARCIA C A.

(KRICK// KRICKER M.

(SLAT// SLATER T.

(DAVI// DAVIS K R.

(ALLE// ALLEN K.

(HOFF// HOFFMAN N.

(HURB// HURBAN P.

Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,

Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,

Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N,

Hurban P;

WPI; 2002-267486/31.

New Arabidopsis thaliana nucleic acid, for identifying homologous

genes, producing compositions that modulate the expression or function

of its encoded protein, and mapping functional regions of a protein -

Claim 1; SEQ ID 615; 44pp; English.

The present invention describes an Arabidopsis thaliana nucleic acid (I)

comprising a sequence capable of hybridising under stringent conditions

to a sequence (SI) selected from any one of the 999 sequences given in

ABL93236 to ABL94234. (I) have insecticide and fungicide activities, and

they can be used as protein expression modulators. (I) can be used in

identifying homologous or related genes, in producing compositions that

modulate the expression or function of their encoded proteins, mapping

functional regions of the proteins, and in studying associated

physiological pathways. (I) can also be used: (1) for the genetic

manipulation of cells, particularly plant cells; (2) in screening assays

of various plant strains to determine the strains that are best capable

of withstanding a particular disease or environmental stress; (3) for

enhancing or inhibiting production of a biosynthetic product in a plant;

(4) as probes in mapping and in diagnosis, in genetic modification and

for screening purposes, to generate additional copies of the nucleic

acids, to generate ribozymes or antisense oligonucleotides, and as

single-stranded DNA probes or as triple-strand forming oligonucleotides;

and (5) for generating genetically modified transgenic organisms.

Note: The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from the

USPTO web site.

Sequence 453 BP; 111 A; 100 C; 132 G; 107 T; 3 other;

Query Match

Best Local Similarity 26.7%; Score 286; DB 24; Length 453;

Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AB013410-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 755 BP; 270 A; 282 C; 78 G; 125 T; 0 other;

Query Match 2.0%; Score 21; DB 24; Length 755;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

438 AAAAAATCCCGACGACGACGAA 458

668 AAAAAATCCCGACGACGACGAA 688

RESULT 6

ABA08516/C

ID ABA08516 standard; cDNA; 453 BP.

ABA08516;

11-JAN-2002 (first entry)

Human Maguin-2 homologue-encoding cDNA, SEQ ID NO:292.

Human; cytokine; cell proliferation; cell differentiation; growth factor; haematopoiesis regulation; tissue growth; immunomodulator; activin; inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis; proliferation; metastasis; cancer; tumour; haematopoietic disorder; myeloid cell disorder; lymphoid cell disorder; asthma; arthritis; chronic inflammatory condition; proliferative retinopathy; atherosclerosis; coronary heart disease; arterial ischaemia; bone disorder; osteoporosis; vascular growth disorder; tissue regeneration; wound healing; infection; immune disorder; cell culture; drug screening; gene therapy; antiinflammatory; antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic; cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial; antifungal; vulnery; antiulcer; ss.

Homo sapiens.

WO200157188-A2.

09-AUG-2001.

05-FEB-2001; 2001WO-US03800.

03-FEB-2000; 2000US-0498914.

27-APR-2000; 2000US-0560875.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-457740/49.

P-PSDB; ABB11272.

Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer.

Claim 1; Page 438; 1963pp; English.

Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may

have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention.

Sequence 453 BP; 109 A; 101 C; 119 G; 124 T; 0 other;

Query Match 1.9%; Score 20; DB 22; Length 453;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 618 CACTGGGCTCAGCAACTTC 637

Db 239 CACTGGGCTCAGCAACTTC 220

RESULT 7

AAC41435

ID AAC41435 standard; DNA; 416 BP.

XX

XX AAC41435;

17-OCT-2000 (first entry)

XX Zea mays DNA fragment SEQ ID NO: 31870.

XX Hybridisation assay; genetic mapping; gene expression control; protein identification; signal transduction pathway; metabolic; pathway; promoter; termination sequence; corn; ss.

XX Zea mays subsp. mays.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 08-APR-1999; 99US-0128234.

XX 16-APR-1999; 99US-0128714.

XX 19-APR-1999; 99US-0129845.

XX 21-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131443.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132484.
PR 04-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139500.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140355.
PR 28-JUN-1999; 99US-0140623.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142320.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147132.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149356.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

```
14-OCT-1999; 99US-0159637.
14-OCT-1999; 99US-0159638.
18-OCT-1999; 99US-0159584.
21-OCT-1999; 99US-0160741.
21-OCT-1999; 99US-0160767.
21-OCT-1999; 99US-0160768.
21-OCT-1999; 99US-0160770.
21-OCT-1999; 99US-0160814.
21-OCT-1999; 99US-0160815.
22-OCT-1999; 99US-0160880.
22-OCT-1999; 99US-0160981.
22-OCT-1999; 99US-0160989.
25-OCT-1999; 99US-0161404.
25-OCT-1999; 99US-0161405.
26-OCT-1999; 99US-0161359.
26-OCT-1999; 99US-0161360.
26-OCT-1999; 99US-0161361.
28-OCT-1999; 99US-0161920.
28-OCT-1999; 99US-0161992.
28-OCT-1999; 99US-0161993.
29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 1.8%; Score 19; DB 21; Length 416;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 761 CGCGCGCGCTGCTCGCC 779
b 57 CGCGCGCGCTGCTCGCC 75

RESULT 8
AC36816
D AAC36816 standard; DNA; 499 BP.
X C
X AAC36816;
X C
X 17-OCT-2000 (first entry)
T T
X Arabidopsis thaliana DNA fragment SEQ ID NO: 15176.
E Arabidopsis thaliana DNA fragment SEQ ID NO: 15176.
W Hybridisation assay; genetic mapping; gene expression control;
W Protein identification; signal transduction pathway;
W Metabolic pathway; promoter; termination sequence; ss.
X Arabidopsis thaliana.
X Arabidopsis thaliana.
N EP1033405-A2.
D 06-SEP-2000.
X X
X 25-FEB-2000; 2000EP-0301439.
F F
R 25-FEB-1999; 99US-0121825.
R 05-MAR-1999; 99US-0123180.
R 09-MAR-1999; 99US-0123548.
R 23-MAR-1999; 99US-0125788.
R 25-MAR-1999; 99US-0126264.
R 29-MAR-1999; 99US-0126785.
R 01-APR-1999; 99US-0127462.
R 06-APR-1999; 99US-0128234.
R 08-APR-1999; 99US-0128714.
R 16-APR-1999; 99US-0129845.
R 19-APR-1999; 99US-0130077.
R 21-APR-1999; 99US-0130449.
R 23-APR-1999; 99US-0130510.
R 23-APR-1999; 99US-0130891.
R 28-APR-1999; 99US-0131449.
R 30-APR-1999; 99US-0132048.
R 30-APR-1999; 99US-0132407.
R 04-MAY-1999; 99US-0132484.
R 05-MAY-1999; 99US-0132485.

106-MAY-1999; 99US-0132486.
106-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
18-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
04-JUN-1999; 99US-0137528.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
20-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
```


CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX Sequence 600 BP; 136 A; 53 C; 216 G; 195 T; 0 other;
 XX Query Match 1.8%; Score 19; DB 24; Length 600;
 XX Best Local Similarity 100.0%; Pred. No. 49;
 XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

YY 441 AATCCGACGACGACGAAC 459
 YY |||||
 YY 238 AATCCGACGACGACGAAC 220

RESULT 10

ABQ43703

ID ABQ43703 standard; DNA; 600 BP.

XX

AC

ABQ43703;

12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 30294.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 drug; side effect; cancer; central nervous system; cardiovascular;
 gastrointestinal; respiratory system; single nucleotide polymorphism;
 SNP; cell differentiation; ds.

Homo sapiens.

WO200218632-A2.

07-MAR-2002.

01-SEP-2001; 2001WO-EP10074.

01-SEP-2000; 2000DE-1043826.

05-SEP-2000; 2000DE-1044543.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K, Guetig D;

WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful
 for diagnosis and prognosis, comprises selective hybridization of
 amplicons from chemically treated DNA

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of
 methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 genomic sample of DNA. The sample is treated chemically to convert
 cytosine (C) but not methylated C, to uracil, then part of the genomic
 DNA that contains the target C is amplified to form a labeled amplicon.
 The amplicon is hybridised to two classes, each with at least one
 member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 and the degree of hybridisation to both classes is determined from the
 label on the amplicon. From the ratio of labels hybridised to the two
 classes of oligomers, the degree of methylation is calculated. The method
 is used: (i) for diagnosis and/or prognosis of side effects of
 therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 of the central nervous, cardiovascular, gastrointestinal and respiratory

CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX Sequence 600 BP; 195 A; 216 C; 136 G; 53 T; 0 other;

XX Query Match 1.8%; Score 19; DB 24; Length 600;

XX Best Local Similarity 100.0%; Pred. No. 49;

XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 441 AATCCGACGACGACGAAC 459

|||

DB 363 AATCCGACGACGACGAAC 381

RESULT 11

ABL99906/C

ID ABL99906 standard; cDNA; 908 BP.

XX

AC

ABL99906;

03-OCT-2002 (first entry)

Human secretory polynucleotide (sptm) 161.

Human; ss; gene; secretory protein; secretory polynucleotides; SPTM;
 SPTM-related disease; somatic gene therapy; germline gene therapy;
 severe combined immunodeficiency; intracellular parasite protection;
 fungal parasite; protozoan parasite; cell proliferative disorder; cancer;
 immune disorder; AIDS; neurological disorder; Parkinson's disease;
 motor neuron disorder; demyelinating disease; multiple sclerosis;
 meningitis; abscess; prion diseases; cerebral palsy;
 neuroskeletal disorder; peripheral nervous system disorder;
 dermatomyositis; polymyositis; myopathy; myasthenia gravis;
 mental disorder; Tourette's syndrome.

Homo sapiens.

WO200220756-A2.

14-MAR-2002.

30-AUG-2001; 2001WO-US27297.

05-SEP-2000; 2000US-229747P.

05-SEP-2000; 2000US-229748P.

05-SEP-2000; 2000US-229749P.

05-SEP-2000; 2000US-229750P.

05-SEP-2000; 2000US-229751P.

05-SEP-2000; 2000US-230016P.

05-SEP-2000; 2000US-230583P.

06-SEP-2000; 2000US-230505P.

06-SEP-2000; 2000US-230514P.

06-SEP-2000; 2000US-230515P.

06-SEP-2000; 2000US-230517P.

06-SEP-2000; 2000US-230518P.

06-SEP-2000; 2000US-230519P.

06-SEP-2000; 2000US-230595P.

06-SEP-2000; 2000US-230596P.

06-SEP-2000; 2000US-230597P.

06-SEP-2000; 2000US-230599P.

06-SEP-2000; 2000US-230610P.

06-SEP-2000; 2000US-230864P.

06-SEP-2000; 2000US-230865P.

06-SEP-2000; 2000US-230988P.

06-SEP-2000; 2000US-230989P.

06-SEP-2000; 2000US-230990P.

07-SEP-2000; 2000US-230896P.

07-SEP-2000; 2000US-230897P.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.8%; Score 19; DB 21; Length 1156;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 CTAAGACCGTCACACGAA 242
|||||
Db 287 CTAAGACCGTCACACGAA 305
|||||

RESULT 13
AAC43180
ID AAC43180 standard; DNA; 1893 BP.
XX
AC AAC43180;
XX
DT 17-OCT-2000 (first entry)
XX
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 38322.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
DD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 25-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131149.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0132856.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

```

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.8%; Score 19; DB 21; Length 1893;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 553 TCCTCGGTTTACGCCAA 571
D 859 TCCTCGGTTTACGCCAA 877

RESULT 14
BL99921/C
D ABL99921 standard; cDNA; 2794 BP.
X
X ABL99921;
X
X 03-OCT-2002 (first entry)
X
X Human secretory polynucleotide (sptm) 176.
X
X Human; ss; gene; secretory protein; secretory polynucleotides; SPTM;
X SPTM-related disease; somatic gene therapy; germline gene therapy;
X severe combined immunodeficiency; intracellular parasite protection;
X fungal parasite; protozoan parasite; cell proliferative disorder; cancer;
X immune disorder; AIDS; neurological disorder; Parkinson's disease;
X motor neuron disorder; demyelinating disease; multiple sclerosis;
X meningitis; abscess; prion diseases; cerebral palsy;
X neuroskeletal disorder; peripheral nervous system disorder;
X dermatomyositis; polymyositis; myopathy; myasthenia gravis;
X mental disorder; Tourette's syndrome.
X
X Homo sapiens.
X
X WO200220756-A2.
X
X 14-MAR-2002.
X
X 30-AUG-2001; 2001WO-US27297.
X
X 05-SEP-2000; 2000US-229747P.
X 05-SEP-2000; 2000US-229748P.
X 05-SEP-2000; 2000US-229749P.
X 05-SEP-2000; 2000US-229750P.
X 05-SEP-2000; 2000US-229751P.
X 05-SEP-2000; 2000US-230016P.
X 05-SEP-2000; 2000US-230583P.
X 06-SEP-2000; 2000US-230505P.
X 06-SEP-2000; 2000US-230514P.

```

```

PR 06-SEP-2000; 2000US-230515P.
PR 06-SEP-2000; 2000US-230517P.
PR 06-SEP-2000; 2000US-230518P.
PR 06-SEP-2000; 2000US-230519P.
PR 06-SEP-2000; 2000US-230595P.
PR 06-SEP-2000; 2000US-230596P.
PR 06-SEP-2000; 2000US-230597P.
PR 06-SEP-2000; 2000US-230599P.
PR 06-SEP-2000; 2000US-230610P.
PR 06-SEP-2000; 2000US-230864P.
PR 06-SEP-2000; 2000US-230865P.
PR 06-SEP-2000; 2000US-230988P.
PR 06-SEP-2000; 2000US-230989P.
PR 06-SEP-2000; 2000US-230990P.
PR 07-SEP-2000; 2000US-230896P.
PR 07-SEP-2000; 2000US-230897P.
PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231832P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
XX Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
XX Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
XX Gerstin BH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
XX Marwaha R, Chen AP, Chang SC, Au AP, Inman RR;
XX
XX WPI; 2002-315658/35.
XX P-PSDB; ABB97925.
XX
XX Polynucleotide sequences encoding human secretory proteins useful for
XX gene therapy of e.g. genetic deficiency disorders, cancers, and
XX diseases caused by intracellular parasites -
XX
XX Claim 1; Page 360-361; 585pp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
XX secretory (SPTM) proteins. The SPTM DNA and amino acid sequences are
XX useful for treating a disease or condition associated with the expression
XX of functional SPTM. The SPTM DNA sequences are useful for somatic or
XX germline gene therapy to correct a genetic deficiency (e.g. severe
XX combined immunodeficiency). The SPTM DNA sequences are also useful in
XX providing protection against intracellular parasites (e.g. fungal
XX parasites and protozoan parasites). The SPTM DNA and protein sequences
XX are also useful for diagnosing cell proliferative disorders, cancer,
XX immune disorders (e.g. AIDS), neurological disorders (e.g. Parkinson's
XX disease), motor neuron disorders, demyelinating diseases (e.g. multiple
XX sclerosis), meningitis, abscesses, prion diseases, cerebral palsy,
XX neuroskeletal disorders, peripheral nervous system disorders,
XX dermatomyositis and polymyositis, myopathy, myasthenia gravis, and mental
XX disorders (e.g. Tourette's syndrome). cDNA sequences ABL99746 - ABL99929
XX represent human secretory polynucleotides of the invention.
XX
XX Sequence 2794 BP; 704 A; 656 C; 549 G; 885 T; 0 other;
XX
XX Query Match 1.8%; Score 19; DB 24; Length 2794;
XX Best Local Similarity 100.0%; Pred. No. 44;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 988 AGAGACTTCTCCCTAGAGA 1006
XX DB 1901 AGAGACTTCTCCCTAGAGA 1883
XX
XX RESULT 15
XX ABL14198
XX ID ABL14198 standard; cDNA; 15987 BP.
XX
XX AC ABL14198;
XX
XX 26-MAR-2002 (first entry)
XX
XX

```


DE : Drosophila melanogaster expressed polynucleotide SEQ ID NO 37076.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
DS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
DR P-PSDB; ABB70095.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 37076; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 15987 BP; 4295 A; 3570 C; 3827 G; 4295 T; 0 other;

Query Match 1.8%; Score 19; DB 23; Length 15987;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 571 ATTGCCACGACGACGCA 589
|||
Db 10662 ATTGCCACGACGACGCA 10680

Search completed: February 2, 2004, 03:13:24
Job time : 317.239 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

us-09-938-842a-1034
Title: US-09-938-842A-1034
Effect score: 1071
Sequence: 1 atggcgacattcagaagct.....cagggtcatcgaaacctga 1071
Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0
22781392 seqs, 12152238056 residues
Word size: 0
Total number of hits satisfying chosen parameters: 45562784
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database: EST:
1: em_estba.*
2: em_esthm.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hcc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hcc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	334	31.2	373	10	BE522768
2	320	29.9	486	9	AV825375
3	244	22.8	378	10	BE523004
4	224	20.9	388	14	Z35068

5	108	10.1	335	14	T88453
6	38	3.5	586	28	BH741132
7	38	3.5	798	28	BH511920
8	35	3.3	656	28	BH450567
9	32	3.0	600	28	BH735681
10	24	2.2	318	9	AM092131
11	24	2.2	481	9	AI775374
12	24	2.2	610	9	AI484214
13	24	2.2	642	9	AM041016
14	24	2.2	670	10	BG129822
15	23	2.1	222	10	BB042901
16	23	2.1	222	10	BB052744
17	23	2.1	222	10	BB052810
18	23	2.1	224	10	BB067092
19	23	2.1	226	10	BB145694
20	23	2.1	232	10	BB118640
21	23	2.1	233	10	BB134002
22	23	2.1	307	10	BB494924
23	23	2.1	381	13	BY506397
24	23	2.1	390	13	BY676112
25	23	2.1	395	13	BY494054
26	23	2.1	415	14	H36511
27	23	2.1	435	13	BY583791
28	23	2.1	441	13	BY436758
29	23	2.1	479	13	BU635806
30	23	2.1	518	9	AI484215
31	21	2.0	499	9	AI795304
32	21	2.0	510	9	AI746274
33	21	2.0	559	9	AI668367
34	21	2.0	567	9	AI665001
35	21	2.0	721	28	BZ001843
36	21	2.0	747	13	BQ865644
37	20	1.9	146	28	BH617024
38	20	1.9	208	28	AZ213090
39	20	1.9	278	28	BZ377781
40	20	1.9	339	28	BH657297
41	20	1.9	371	9	AA645472
42	20	1.9	377	10	BF649315
43	20	1.9	385	9	AI506347
44	20	1.9	400	14	CB699695
45	20	1.9	427	14	CB795370

ALIGNMENTS

RESULT 1	BE522768	373 bp	mRNA	linear	EST 19-MAR-2001
LOCUS	M28F68STM	Arabidopsis	developing seed	Arabidopsis thaliana	cdna
DEFINITION	Clone M28F8 5', mRNA sequence.				
ACCESSION	BE522768				
VERSION	BE522768.1	GI:9780746			
KEYWORDS	EST.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	White, J.A., Todd, J.J., Newman, T., Focks, N., Girke, T., Martinez de Irla, O., Javorski, J.G., Chirgove, J., and Benning, C. 1999. A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil Plant Physiol. 124 (4), 1582-1594 (2000)				
AUTHORS	White, J.A., Todd, J.J., Newman, T., Focks, N., Girke, T., Martinez de Irla, O., Javorski, J.G., Chirgove, J., and Benning, C.				
TITLE	A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil				
JOURNAL	Plant Physiol. 124 (4), 1582-1594 (2000)				
MEDLINE	20567808				
PUBMED	11115876				
COMMENT	Contact: Benning, C Dept. of Biochemistry & Molecular Biology Michigan State University 224 Biochemistry, Michigan State University, East Lansing, MI 48824 , USA Tel: 517 355 1609				

Fax: 517 353 9334

Email: benning@osu.edu

Michigan State University DNA Sequencing Facility Arabidopsis

Biological Resource Center, The Ohio State University, 309 Botany &

Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:

6142920603 TEL: 6142929371.

Location/Qualifiers

FEATURES

source

1. 373
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="M28P8"
 /tissue_type="seed"
 /dev_stage="5-13 days after flowering"
 /lab_host="E.coli"
 /clone_lib="Arabidopsis developing seed"
 /notes="Organ: Developing seed; Vector: pBluescript SK-;
 Site1: EcoRI; Site2: XhoII"
 73 a 118 c 82 g 100 t

BASE COUNT

ORIGIN

Query Match 31.2%; Score 334; DB 10; Length 373;

Best Local Similarity 100.0%; Pred. No. 6.9e-174;

Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 536 GCGACGCGTTTCAGCTTCCTCGGTTTGTAGTCCAAATGCGCAGCAGCAGCAAGATCCAC 595

Db 12 GCGACGCGTTTCAGCTTCCTCGGTTTGTAGTCCAAATGCGCAGCAGCAGCAAGATCCAC 71

QY 596 CTCGCAAGCTCTGGCATCTCACTGTGCTGCTCAGCACTTCGCGCAGGAAGTATC 655

Db 72 CTCGCAAGCTCTGGCATCTCACTGTGCTGCTCAGCACTTCGCGCAGGAAGTATC 131

QY 656 CGATGTGGGTATTCCATCAAAACCAATGATTCGACGGTTCGAGCTTCTCTTGAATC 715

Db 132 CGATGTGGGTATTCCATCAAAACCAATGATTCGACGGTTCGAGCTTCTCTTGAATC 191

QY 716 CACAATCGTGTGCTCGTCAATCAGCCCTAGTATTAGCTTTCCGCGCGGCTGCTT 775

Db 192 CACAATCGTGTGCTCGTCAATCAGCCCTAGTATTAGCTTTCCGCGCGGCTGCTT 251

QY 776 CGCGCTGCTTACGTGCGCGCTTTCAACAGCGCTTCCAGTGGCTAGACCACTCTTT 835

Db 252 CGCGCTGCTTACGTGCGCGCTTTCAACAGCGCTTCCAGTGGCTAGACCACTCTTT 311

QY 836 TACAAGTTGTTCAGCAGCGGCTTTGTATCCGT 869

Db 312 TACAAGTTGTTCAGCAGCGGCTTTGTATCCGT 345

RESULT 2

AV825375

LOCUS AV825375 RAFL7 Arabidopsis thaliana cDNA clone RAFL07-08-P04 5',

DEFINITION AV825375 mRNA sequence. EST 01-APR-2002

ACCESSION AV825375

VERSION AV825375.1 GI:19867435

KEYWORDS Arabidopsis thaliana (thale cress)

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 486)

Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,

Ono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,

Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.

and Shinozaki, K.

Large scale analysis of Arabidopsis full-length cDNA (2002b)

Unpublished

CONTACT: Morioaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: mae@rc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially

as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

and XhoI was ligated to modified Lambda PLC-1 vector (Carninci et

al., submitted for publication) digested with BamHI and SalI. This

clone is in a modified pBluescript vector. Please visit our web

site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further

details.

Location/Qualifiers

FEATURES

source

1. 486
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL07-08-P04"
 /dev_stage="rosette plants"
 /lab_host="DH10B"
 /clone_lib="RAFL7"
 /notes="Site 1: BamHI; Site 2: SalI; subjected to
 cold-treated (1, 2, 5, 10, 24 hr)"
 131 a 118 c 117 g 117 t 3 others

BASE COUNT

ORIGIN

Query Match 29.9%; Score 320; DB 9; Length 486;

Best Local Similarity 100.0%; Pred. No. 4.5e-166;

Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGCAATTCAGAGCTTGAAGAGTTGAGGAGTTCAGGCAAGATCAAACTCTAAGAGCGGTT 60

Db 102 ATGCGCAATTCAGAGCTTGAAGAGTTGAGGAGTTCAGGCAAGATCAAACTCTAAGAGCGGTT 161

QY 61 GATCTAACCATCATCAACGGCTCAGAAAGCTCGAAACTTCAGACCTTCCAAAGTAAT 120

Db 162 GATCTAACCATCATCAACGGCTCAGAAAGCTCGAAACTTCAGACCTTCCAAAGTAAT 221

QY 121 CCCACAGTGTCTCGAGCCCAAGGGGAGCGGTGATCGCGTTCGATGCTTTCAATGCTTTTA 180

Db 222 CCCACAGTGTCTCGAGCCCAAGGGGAGCGGTGATCGCGTTCGATGCTTTCAATGCTTTTA 281

QY 181 GCTCCACCGCTTCGACAGGACCACTTGAAGAGAGCTTCGACTAAAGACCGTCAACG 240

Db 282 GCTCCACCGCTTCGACAGGACCACTTGAAGAGAGCTTCGACTAAAGACCGTCAACG 341

QY 241 AAGTTGAAGAGAGGAGAGATACGGATCCCTGCCACGTTGCGGCTAGGATTTT 300

Db 342 AAGTTGAAGAGAGGAGAGATACGGATCCCTGCCACGTTGCGGCTAGGATTTT 401

QY 301 CAATTAACCTGAGAGTTAGG 320

Db 402 CAATTAACCTGAGAGTTAGG 421

RESULT 3

BE523004

LOCUS BE523004

DEFINITION BE523004 378 bp mRNA linear EST 19-MAR-2001

ACCESSION BE523004

VERSION BE523004.1 GI:9780982

KEYWORDS Arabidopsis thaliana (thale cress)

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 378)

White, J.A., Todd, J., Newman, T., Pocks, N., Girke, T., Martinez de

Ibarra, O., Jaworski, J.G., Ohlrogge, J., and Benning, C.

A new set of Arabidopsis expressed sequence tags from developing

seeds. The metabolic pathway from carbohydrates to seed oil

Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142923371.

```

SE COUNT
109 a_97 c      81 q_92 t
Site 1: ECORI; Site 2: XhoI;
/clone_lib="Arabidopsis developing seed"
/notes="Organ: Developing seed; Vector: pBluescript SK-;
/dev stages="5-13 days after flowering"
/lab_host="E.coli"
/clone="M3.1B6"
/tissue_type="seed"

```

1 ATGGCGCAATTGAGAGCTTGAAGAAGTTGCAGGCAGGCAAGATCAAACTCTAAGAGCGTT 60
103 ATGGCGCAATTGAGAGCTTGAAGAAGTTGCAGGCAGGCAAGATCAAACTCTAAGAGCGTT 162

121 CCCACAGTGAGTCTCGAGGCCCAAGCGCGGATGCGGTCGTTTTCAATGCTCTTA 180
|||
223 CCCACAGTGAGTCTCGAGGCCCAAGCGCGGATGCGGTCGTTTTCAATGCTCTTA 282

241 AAGG 244
|||
343 AAGG 346

Accession	Gene	Size	Type	Source	Accession
35068	OCUS	388 bp	mRNA	linear	EST 18-JUL-1994
Z35068	AT5G3707				
	Glf-SeedA				
	Arabidopsis				
	thaliana				
	CDNA clone				
	YAY344				
	5', mRNA				
	sequence				

SOURCE	ORGANISM
Arabidopsis thaliana (thale cress)	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
Arabidopsis thaliana	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
Arabidopsis thaliana	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

TITLE The Arabidopsis thaliana transcribed genome: the GDR cDNA program
JOURNAL Unpublished

/organism="Arabidopsis thaliana"

/mol_type="rRNA"

/strain="var columbia"

/db_xref="taxon:3702"

/clone="156J17"

/clone_lib="Lambda-PR2"

/note="Vector: lambda Zip-Lox; Site 1: Sal; Site 2: Not; Lambda PR2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRU's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."

BASE COUNT 96 a 75 c 82 g 69 t 13 others

ORIGIN

Query Match 10.1%; Score 108; DB 14; Length 335;

Best Local Similarity 100.0%; Pred. No. 2.7e-48; Length 335;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 42 TCAACTCTAAGAGCGGTGATCAACCATCATCAACGGCGTCAGAAAGCTCGAAACTTC 101

2b 40 TCAACTCTAAGAGCGGTGATCAACCATCATCAACGGCGTCAGAAAGCTCGAAACTTC 99

2Y 102 AAGACCTTCCAAAGTAATCCACAGTGTCTGAGCCCAAGCGGA 149

2b 100 AAGACCTTCCAAAGTAATCCACAGTGTCTGAGCCCAAGCGGA 147

RESULT 6

3B741132/c

LOCUS BH741132 586 bp DNA linear GSS 25-FEB-2002
DEFINITION gt27g10.g1 BoBuds01 Brassica oleracea genomic clone gt27g10 5',
genomic survey sequence.

ACCESSION BH741132

VERSION BH741132.1 GI:18875745

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 586)

AUTHORS Katari, M., O'Shaughnessy, A., Palmer, L., Bahret, A., Baker, J., Balija V., Cunniss, D.M., Katzenberger, F., King, L., Kirchoff, K., Kuit, K., Miller, B., Muller, S., Nascimento, L., Preston, R., Santos, L., Shah, R., Zutavern, T., Dedhia, N., Rabinowicz, P.D. and McCombie, W.R.

Whole Genome Shotgun Reads from Brassica oleracea (2002b)

Unpublished

CONTACT: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: gt27 row: g column: 10

Seq primer: -21UnivRev

Class: shotgun

High quality sequence stop: 586.

Location/Qualifiers

source

1. 586

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/db_xref="taxon:3712"

/clone="gt27g10"

/clone_lib="BoBuds01"

/note="Vector: M13 for .x reads, pBluescript for .b and .g reads; Site 1: EcoRV; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep

using Brassica oleracea TC1000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA provided by Pablo Rabinowicz (CSHL) and shotgun library prepared in McCombie Lab."

BASE COUNT 162 a 120 c 131 g 173 t

ORIGIN

Query Match 3.5%; Score 38; DB 28; Length 586;

Best Local Similarity 100.0%; Pred. No. 2.6e-09;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 AGGATACGGATGCTGCCACGTGTGGCGCTAGGATTTT 299

Db 41 AGGATACGGATGCTGCCACGTGTGGCGCTAGGATTTT 4

RESULT 7

BH511920/c

LOCUS BH511920 798 bp DNA linear GSS 13-DEC-2001
DEFINITION BOHGA14TF BOHG Brassica oleracea genomic clone BOHGA14, genomic
survey sequence.

ACCESSION BH511920

VERSION BH511920.1 GI:17720010

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 798)

AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Other GSSs: BOHGA14TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

source

1. 798

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TC1000DH3"

/db_xref="taxon:3712"

/clone="BOHGA14"

/clone_lib="BOHG"

/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHOS1 using BstXI linkers"

BASE COUNT 161 a 210 c 225 g 202 t

ORIGIN

Query Match 3.5%; Score 38; DB 28; Length 798;

Best Local Similarity 100.0%; Pred. No. 2.8e-09;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 AGGATACGGATGCTGCCACGTGTGGCGCTAGGATTTT 299

Db 720 AGGATACGGATGCTGCCACGTGTGGCGCTAGGATTTT 683

RESULT 8

BH450567/c

LOCUS BH450567 656 bp DNA linear GSS 12-DEC-2001
DEFINITION BOGDQ41TR BOGD Brassica oleracea genomic clone BOGDQ41, genomic
survey sequence.

ACCESSION BH450567

VERSION BH450567.1 GI:17636278

KEYWORDS GSS.

SOURCE
ORGANISM
Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 656)

REFERENCE
TOWN, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOGDQ41TF
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: Sheared ends.

FEATURES
source
Location/Qualifiers
1..656
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGDQ41"
/clone_lib="BOGD"
/note="Vector: pBstXI; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pBstXI using BstXI linkers"
142 a 167 c 190 g 157 t

Query Match 3.3%; Score 35; DB 28; Length 656;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 382 CCGCCACGGGTAGCGGAACGGTTCGCCCATCGC 416
D 563 CCGCCACGGGTAGCGGAACGGTTCGCCCATCGC 529

RESULT 9
H735681
OCUS
EFINITION
BOHTG58TR BO 2.3 KB Brassica oleracea genomic clone BOHTG58,
genomic survey sequence.

CESSION
H735681
ERSON
H735681
EYWORDS
GSS.
SOURCE
Brassica oleracea
Brassica oleracea

ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 600)

REFERENCE
TOWN, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BOHTG58TF
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: Sheared ends.

FEATURES
source
Location/Qualifiers
1..600
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"

/db_xref="taxon:3712"
/clone="BOHTG58"
/clone_lib="BO 2.3 KB"
/note="Vector: pBstXI; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pBstXI using BstXI linkers"
159 a 126 c 160 g 155 t

BASE COUNT
ORIGIN

Query Match 3.0%; Score 32; DB 28; Length 600;
Best Local Similarity 100.0%; Pred. No. 5.6e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 ACTCGAGAGTTAGGTCCAAATCCGACGGCGA 338
DB 119 ACTCGAGAGTTAGGTCCAAATCCGACGGCGA 150

RESULT 10

AW092131

LOCUS

DEFINITION

AW092131

VERSION

AW092131.1

KEYWORDS

SOURCE

ORGANISM

Lycopersicon esculentum (tomato)

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 318)

REFERENCE

AUTHORS

D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,
Romling, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Niemman, W.,
Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni,
J.

Generation of ESTs from tomato leaf tissue

Unpublished

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

source

1..318

/organism="Lycopersicon esculentum"

/mol_type="mRNA"

/cultivar="Rio Grande PtoR"

/db_xref="taxon:4081"

/clone="cLET17P7"

/tissue_type="leaf"

/dev_stage="4-6 week old plants"

/lab_host="XLa-Blue MRF"

/clone_lib="tomato mixed elicitor, BTI"

XhoI; cLET - inoculated with a variety of disease response

elicitors. Plants exposed to 2,6 dichloroisocoumaric

acid, BTH, jasmonic acid, ethylene, fenthion, EIX,

okadaic acid, or systemin prior to tissue harvest. EORI

site was destroyed during cloning."

58 a 78 c 82 g 100 t

BASE COUNT

ORIGIN

Query Match 2.2%; Score 24; DB 9; Length 318;

Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 CCGCCGCCCTGCTTCGCCGCTGCT 784

DB 36 CCGCCGCCCTGCTTCGCCGCTGCT 59

RESULT 11

AI775374
 LOCUS
 DEFINITION
 EST256474 tomato resistant, Cornell Lycopersicon esculentum cDNA
 clone cLER15G18, mRNA sequence.
 ACCESSION
 AI775374
 VERSION
 AI775374.1
 SOURCE
 AI775374.1 GI:5273415
 ORGANISM
 Lycopersicon esculentum (tomato)
 ORGANISM
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 481)
 REFERENCE
 D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
 Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman
 C.L., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
 Giovannoni, J.J. and Martin, G.B.
 Generation of ESTs from Pseudomonas resistant tomato
 Unpublished
 TITLE
 Unpublished
 JOURNAL
 COMMENT
 Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.
 Location/Qualifiers
 1..481
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
 /db_xref="taxon:4081"
 /clone="cLER15G18"
 /tissue_type="leaf"
 /dev_stage="4-week old"
 /lab_host="SOLR"
 /clone_lib="tomato resistant, Cornell"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; cLER - Tomato Pseudomonas Resistant EST Library.
 Directionally cloned cDNAs inserted into pBluescript SK(-
) at 5' end with EcoRI and 3' end with XhoI site."
 BASE COUNT 109 a 114 c 110 g 148 t
 ORIGIN
 Query Match 2.2%; Score 24; DB 9; Length 481;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 2y 761 CCGCCCGCTGCTTCGCGTCTGT 784
 80 CCGCCCGCTGCTTCGCGTCTGT 103
 EST.
 SOURCE
 AI484214
 LOCUS
 DEFINITION
 EST249484 tomato resistant, Cornell Lycopersicon esculentum cDNA
 clone cLER1A9, mRNA sequence.
 ACCESSION
 AI484214
 VERSION
 AI484214.1
 SOURCE
 AI484214.1 GI:4379585
 ORGANISM
 Lycopersicon esculentum (tomato)
 ORGANISM
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 610)
 REFERENCE
 D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
 Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman
 C.L., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
 Giovannoni, J.J. and Martin, G.B.
 Generation of ESTs from Pseudomonas resistant tomato
 Unpublished
 TITLE
 Unpublished
 JOURNAL
 COMMENT
 Other_ESTs: EST249485

Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>.
 Location/Qualifiers
 1..610
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
 /db_xref="taxon:4081"
 /clone="cLER1A9"
 /tissue_type="leaf"
 /dev_stage="4-week old"
 /lab_host="SOLR"
 /clone_lib="tomato resistant, Cornell"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; cLER - Tomato Pseudomonas Resistant EST Library.
 Directionally cloned cDNAs inserted into pBluescript SK(-
) at 5' end with EcoRI and 3' end with XhoI site."
 BASE COUNT 194 a 138 c 128 g 150 t
 ORIGIN
 Query Match 2.2%; Score 24; DB 9; Length 610;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 761 CCGCCCGCTGCTTCGCGTCTGT 784
 Db 602 CCGCCCGCTGCTTCGCGTCTGT 579
 EST.
 SOURCE
 AW041016
 LOCUS
 DEFINITION
 EST23880 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
 clone cLET621, mRNA sequence.
 ACCESSION
 AW041016
 VERSION
 AW041016.1
 SOURCE
 AW041016.1 GI:5899770
 ORGANISM
 Lycopersicon esculentum (tomato)
 ORGANISM
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 642)
 REFERENCE
 D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,
 Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W.,
 Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni
 J.J.
 Generation of ESTs from tomato leaf tissue
 Unpublished
 TITLE
 Unpublished
 JOURNAL
 COMMENT
 Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.
 Location/Qualifiers
 1..642
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="Rio Grande PtoR"
 /db_xref="taxon:4081"
 /clone="cLET621"
 /tissue_type="leaf"
 /dev_stage="4-6 week old plants"
 /lab_host="X11-Blue MRF"
 /clone_lib="tomato mixed elicitor, BTI"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; cLER - inoculated with a variety of disease response
 elicitors. Plants exposed to 2,6 dichloroisonicotinic
 acid, BTH, jasmonic acid, ethylene, fenthion, EIX,

ORIGIN

Query Match 2.1%; Score 23; DB 10; Length 222;
 Best Local Similarity 100.0%; Pred. No. 0.41; Mismatches 0; Indels 0; Gaps 0;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 CAAGTAAATCCACAGTGTCT 134
 |||||
 Db 149 CAAGTAAATCCACAGTGTCT 171

Search completed: February 2, 2004, 05:01:17
 Job time : 2490.95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model,
un on: February 2, 2004, 02:56:51 ; Search time 71.3221 Seconds
(without alignments)
6627.983 Million cell updates/sec

title: US-09-938-842A-1034
effect score: 1071
sequence: 1 atggcgacaattcagaagct.....cacggtcatcgaaacctga 1071

coring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

searched: 569978 seqs, 220691566 residues

ord size : 0

total number of hits satisfying chosen parameters: 1139956

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Listing first 45 summaries

atabase : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	19	1.8	34094	US-09-292-034-1	Sequence 1, Appli
2	18	1.7	162450	US-09-345-882-1	Sequence 1, Appli
3	18	1.7	1664976	US-08-916-421B-1	Sequence 1, Appli
4	17	1.6	447	US-09-328-352-186	Sequence 186, App
5	17	1.6	516	US-08-510-878-2	Sequence 2, Appli
6	17	1.6	748	US-08-510-878-3	Sequence 3, Appli
7	17	1.6	867	US-09-252-931A-6878	Sequence 6878, Ap
8	17	1.6	882	US-09-252-931A-6792	Sequence 6792, Ap
9	17	1.6	1353	US-08-611-280-1	Sequence 1, Appli
10	17	1.6	1353	US-09-195-940-1	Sequence 1, Appli
11	17	1.6	1353	US-09-562-466-1	Sequence 1, Appli
12	17	1.6	1960	US-09-562-466-1	Sequence 42, Appli
13	17	1.6	3758	US-09-323-477-1	Sequence 1, Appli
14	17	1.6	12537	US-08-611-280-4	Sequence 4, Appli
15	17	1.6	12537	US-09-195-940-4	Sequence 4, Appli
16	17	1.6	12537	US-09-562-466-4	Sequence 4, Appli
17	17	1.6	1684976	US-08-916-421B-1	Sequence 1, Appli
18	17	1.6	4403765	US-09-103-840A-2	Sequence 2, Appli
19	17	1.6	4411529	US-09-103-840A-1	Sequence 1, Appli
20	16	1.5	268	US-08-444-818-21	Sequence 21, Appli
21	16	1.5	306	US-08-411-913-8	Sequence 8, Appli
22	16	1.5	307	US-08-444-818-143	Sequence 143, App
23	16	1.5	438	US-09-255-000-8	Sequence 8, Appli
24	16	1.5	477	US-07-853-985A-7	Sequence 7, Appli
25	16	1.5	477	US-07-681-703B-7	Sequence 7, Appli
26	16	1.5	477	US-08-184-236-7	Sequence 7, Appli
27	16	1.5	477	US-08-407-410B-7	Sequence 7, Appli

C 28 16 1.5 477 2 US-08-485-500-7 Sequence 7, Appli
C 29 16 1.5 477 5 PCT-US91-02370-7 Sequence 7, Appli
C 30 16 1.5 477 5 PCT-US94-04174-7 Sequence 7, Appli
C 31 16 1.5 489 4 US-09-328-352-1768 Sequence 1768, Ap
C 32 16 1.5 493 4 US-09-364-206-5 Sequence 5, Appli
C 33 16 1.5 557 4 US-09-364-206-6 Sequence 6, Appli
C 34 16 1.5 558 1 US-07-853-985A-9 Sequence 9, Appli
C 35 16 1.5 558 1 US-07-681-703B-9 Sequence 9, Appli
C 36 16 1.5 558 1 US-08-184-236-9 Sequence 9, Appli
C 37 16 1.5 558 2 US-08-407-410B-9 Sequence 9, Appli
C 38 16 1.5 558 2 US-08-485-500-9 Sequence 9, Appli
C 39 16 1.5 558 5 PCT-US91-02370-9 Sequence 9, Appli
C 40 16 1.5 558 5 PCT-US94-04174-9 Sequence 9, Appli
C 41 16 1.5 701 3 US-08-998-416-908 Sequence 908, App
C 42 16 1.5 721 4 US-09-205-258-141 Sequence 141, App
C 43 16 1.5 800 3 US-08-444-818-31 Sequence 31, Appli
C 44 16 1.5 816 1 US-08-350-884-83 Sequence 83, Appli
C 45 16 1.5 816 1 US-08-440-548-83 Sequence 83, Appli

ALIGNMENTS

RESULT 1
US-09-292-034-1
; Sequence 1, Application US/09292034
; Patent No. 6492343
; GENERAL INFORMATION:
; APPLICANT: Reddy, P. Seshidhar
; APPLICANT: Bablu, Suresh
; APPLICANT: Tikoo, Suresh
; TITLE OF INVENTION: PORCINE ADENOVIRUS TYPE 3 GENOME
; FILE REFERENCE: 293102002400
; CURRENT APPLICATION NUMBER: US/09/292,034
; CURRENT FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 34094
; TYPE: DNA
; ORGANISM: Porcine Adenovirus Type 3
; FEATURE:
US-09-292-034-1

Query Match 1.8%; Score 19; DB 4; Length 34094;
Best Local Similarity 100.0%; Pred.No.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 850 AGCAGCGGCTTTGTATCCG 868
Db 21527 AGCAGCGGCTTTGTATCCG 21545

RESULT 2
US-09-345-882-1
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 162450
; TYPE: DNA

```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-124-273 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 88073
OTHER INFORMATION: 5-127-261 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 105940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
```

```

NAME/KEY: allele
LOCATION: 98094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:

```

```
Query Match      1.7%; Score 18; DB 4; Length 162450;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

y
695' TCGAGCTTTCTTCTTGA 712
|||
b
35413 TCGAGCTTTCTTCTTGA 35430

RESULT 3
 S-08-916-421B-1/c
 Sequence 1, Application US/08916421B
 Patent No. 6503729
 GENERAL INFORMATION:
 APPLICANT: Bult et al.
 TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
 Patent No. 6503729
 TITLE OF INVENTION: jannaschii
 FILE REFERENCE: PB275
 CURRENT APPLICATION NUMBER: US/08/916,421B
 CURRENT FILING DATE: 1997-08-22
 PRIOR APPLICATION NUMBER: US 60/024,428
 PRIOR FILING DATE: 1996-08-22
 NUMBER OF SEQ ID NOS: 3
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 1
 LENGTH: 1664976
 TYPE: DNA
 ORGANISM: Methanococcus jannaschii
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (2822)..(28222)
 OTHER INFORMATION: n equals a, t, c, or g
 NAME/KEY: misc feature
 LOCATION: (28257)..(28258)
 OTHER INFORMATION: n equals a, t, c, or g

LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674335)..(674335)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 1.7%; Score 18; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 CAATGCTTTTACCTCCAC 187
DB 1522815 CAATGCTTTTACCTCCAC 1522798

RESULT 4

US-09-328-352-186
Sequence 186, Application US/09328352
Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: BAUMANNI FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 186
LENGTH: 447
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-186

Query Match 1.6%; Score 17; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 47;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 TTCATCGTCAATTGCAA 967
DB 326 TTCATCGTCAATTGCAA 342

RESULT 5

US-08-510-878-2
Sequence 2, Application US/08510878
Patent No. 5776771

GENERAL INFORMATION:

APPLICANT: Yu, Fujio
APPLICANT: Kato, Mami
TITLE OF INVENTION: A KANAMYCIN RESISTANCE GENE DERIVED FROM
TITLE OF INVENTION: MICROORGANISMS OF THE GENUS RHODOCOCUS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,878
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weiner, Marc S
REGISTRATION NUMBER: 32,181
REFERENCE/DOCKET NUMBER: 1254-121
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
S-08-510-878-2

Query Match 1.6%; Score 17; DB 1; Length 516;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 439 AAAATCCCGACGACGAC 455
b 120 AAAATCCCGACGACGAC 136

RESULT 6

S-08-510-878-3
Sequence 3, Application US/08510878
Patent No. 5776771

GENERAL INFORMATION:

APPLICANT: Yu, Fujio
APPLICANT: Kato, Mami
TITLE OF INVENTION: A KANAMYCIN RESISTANCE GENE DERIVED FROM
TITLE OF INVENTION: MICROORGANISMS OF THE GENUS RHODOCOCUS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,878
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weiner, Marc S
REGISTRATION NUMBER: 32,181
REFERENCE/DOCKET NUMBER: 1254-121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 748 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
S-08-510-878-3

Query Match 1.6%; Score 17; DB 1; Length 748;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 439 AAAATCCCGACGACGAC 455
b 284 AAAATCCCGACGACGAC 300

RESULT 7

US-09-252-991A-6878
Sequence 6878, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6878
LENGTH: 867
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6878

Query Match 1.6%; Score 17; DB 4; Length 867;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 CGCGGATTATAGCCGCC 387
b 50 CGCGGATTATAGCCGCC 66

RESULT 8

US-09-252-991A-6792/c
Sequence 6792, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6792
LENGTH: 882
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6792

Query Match 1.6%; Score 17; DB 4; Length 882;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 CGCGGATTATAGCCGCC 387
b 82 CGCGGATTATAGCCGCC 66

RESULT 9

US-08-611-280-1/c
Sequence 1, Application US/08611280
Patent No. 5891666

GENERAL INFORMATION:

APPLICANT: Matsuyama, Toshifumi
APPLICANT: Grossman, Alex
APPLICANT: Richardson, Christopher D.
TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES

NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Canada Inc.
STREET: 6733 Mississauga Road, Suite 303
CITY: Mississauga
STATE: Ontario
COUNTRY: Canada
ZIP: L5N 6J8
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,280
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-338A
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1353 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-611-280-1

Query Match 1.6%; Score 17; DB 2; Length 1353;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 GTGGCTCAGCAACTTCT 638
DB 930 GTGGCTCAGCAACTTCT 914

RESULT 10
US-09-195-940-1/c
; Sequence 1, Application US/09195940
; Patent No. 6258935
; GENERAL INFORMATION:
; APPLICANT: Matsuyama, Toshifumi
; APPLICANT: Grossman, Alex
; APPLICANT: Richardson, Christopher D.
; TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Canada Inc.
; STREET: 6733 Mississauga Road, Suite 303
; CITY: Mississauga
; STATE: Ontario
; COUNTRY: Canada
; ZIP: L5N 6J8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,940
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/611,280
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-338A
; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1353 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-195-940-1

Query Match 1.6%; Score 17; DB 3; Length 1353;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 GTGGCTCAGCAACTTCT 638
DB 930 GTGGCTCAGCAACTTCT 914

RESULT 11
US-09-562-466-1/c
; Sequence 1, Application US/09562466
; Patent No. 6369202
; GENERAL INFORMATION:
; APPLICANT: Matsuyama, Toshifumi
; APPLICANT: Grossman, Alex
; APPLICANT: Richardson, Christopher D.
; TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Canada Inc.
; STREET: 6733 Mississauga Road, Suite 303
; CITY: Mississauga
; STATE: Ontario
; COUNTRY: Canada
; ZIP: L5N 6J8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/562,466
; FILING DATE: 01-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/195,940
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-338A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1353 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-562-466-1

Query Match 1.6%; Score 17; DB 4; Length 1353;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 GTGGCTCAGCAACTTCT 638
DB 930 GTGGCTCAGCAACTTCT 914

RESULT 12
US-09-553-867A-42/c
; Sequence 42, Application US/09553867A
; Patent No. 6476188

GENERAL INFORMATION:

APPLICANT: Young, Michael W
APPLICANT: Kloss, Brian
APPLICANT: Blau, Justin
APPLICANT: Price, Jeffrey
APPLICANT: Takahashi, Joseph S.
APPLICANT: Philip, Lowrey L.
TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE THEREOF
FILE REFERENCE: 600-1-22INC
CURRENT APPLICATION NUMBER: US/09/553.867A
PRIORITY FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: 60/090,068
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 09/335,983
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 42
LENGTH: 1960
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: Primer.
S-09-553-867A-42

Query Match 1.6%; Score 17; DB 4; Length 1960;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 761 CCGCGCGCGTGTTCG 777
D 101 CCGCGCGCGTGTTCG 85

RESULT 13

S-08-323-477-1
Sequence 1, Application US/08323477
Patent No. 6086896
GENERAL INFORMATION:
APPLICANT: Sparling, P. Frederick
APPLICANT: Thompson, Stuart
TITLE OF INVENTION: ANTIGENIC IRON REPRESSIBLE PROTEINS FROM
TITLE OF INVENTION: N. MENINGITIDIS RELATED TO THE HEMOLYSIN FAMILY OF TOXINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: ImClone Systems Incorporated
STREET: 180 Varick Street
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323.477
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/920.963
FILING DATE: 28-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Feit, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: SPA-2-2P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-645-1405
TELEFAX: 212-645-2054
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3758 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Neisseria meningitidis
FEATURE:
NAME/KEY: CDS
LOCATION: 298..3645
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 298..3642
US-08-323-477-1

Query Match 1.6%; Score 17; DB 3; Length 3758;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1003 GAGATATACGAGAAACA 1019
DB 590 GAGATATACGAGAAACA 606

RESULT 14

US-08-611-280-4/c
Sequence 4, Application US/08611280
Patent No. 5891666
GENERAL INFORMATION:
APPLICANT: Matsuyama, Toshifumi
APPLICANT: Grossman, Alex
APPLICANT: Richardson, Christopher D.
TITLE OF INVENTION: NOVEL GENES ENCODING LSIRP POLYPEPTIDES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Canada Inc.
STREET: 6733 Mississauga Road, Suite 303
CITY: Mississauga
STATE: Ontario
COUNTRY: Canada
ZIP: L5N 6J8
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611.280
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-338A
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 12537 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-611-280-4

Query Match 1.6%; Score 17; DB 2; Length 12537;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 GTGGCTCAGCACTTCT 638
DB 8349 GTGGCTCAGCACTTCT 8333

RESULT 15
 US-09-195-940-4/c
 ; Sequence 4, Application US/09195940
 ; Patent No. 6258935
 ; GENERAL INFORMATION:
 ; APPLICANT: Matsuyama, Toshifumi
 ; APPLICANT: Grossman, Alex
 ; APPLICANT: Richardson, Christopher D.
 ; TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amgen Canada Inc.
 ; STREET: 6733 Mississauga Road, Suite 303
 ; CITY: Mississauga
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: L5N 6J8
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/195,940
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/611,280
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Olecki, Nancy A.
 ; REGISTRATION NUMBER: 34,688
 ; REFERENCE/DOCKET NUMBER: A-338A
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 12537 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; JS-09-195-940-4

Query Match 1.6%; Score 17; DB 3; Length 12537;
 Best Local Similarity 100.0%; Pred No. 44;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 622 GTGGCTCAGCACTTCT 638
 |||||
 Db 8349 GTGGCTCAGCACTTCT 8333

Search completed: February 2, 2004, 07:58:50
 Job time : 89.3221 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

un on: February 2, 2004, 03:00:32 ; Search time 801.301 Seconds
(Without alignments)
4871.226 Million cell updates/sec

title: US-09-938-842A-1034
erfect score: 1071
equence: 1 atggcgacaattcagaagct.....cacggtcatgaaccactga 1071

coring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

searched: 2434939 seqs, 1822278265 residues

ord size : 0

otal number of hits satisfying chosen parameters: 4869878

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Listing first 45 summaries

atabase : Published Applications NA:**

- 1: /cgm2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgm2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgm2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgm2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgm2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgm2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgm2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgm2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgm2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgm2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgm2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgm2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgm2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 14: /cgm2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgm2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgm2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 17: /cgm2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 18: /cgm2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1071	100.0	1071	10	US-09-938-842A-1034
2	1071	100.0	1071	12	US-09-938-842A-1034
3	289	27.0	460	10	US-09-938-842A-1034
4	286	26.7	453	9	US-09-938-842A-1034
5	185	17.3	185	9	US-09-770-444-615
6	23	2.1	1604	15	US-10-295-403-147
7	20	1.9	1263	12	US-10-369-493-37704
8	19	1.8	34094	13	US-09-963-038A-1
9	19	1.8	34094	15	US-10-139-550-1
10	19	1.8	34094	16	US-10-245-603A-1
11	18	1.7	277	9	US-09-294-093B-4716
12	18	1.7	364	11	US-09-918-995-30316
13	18	1.7	377	11	US-09-918-995-34549
14	18	1.7	424	9	US-09-777-564-1604
15	18	1.7	424	15	US-10-015-219-1604

c 16	1.7	432	10	US-09-983-965-5485	Sequence 5485, Ap
c 17	1.7	475	9	US-09-770-444-64	Sequence 64, Appl
c 18	1.7	572	13	US-10-027-632-216038	Sequence 216038,
c 19	1.7	572	14	US-10-027-632-216038	Sequence 216038,
c 20	1.7	526	9	US-09-770-149-739	Sequence 739, App
c 21	1.7	1119	12	US-10-369-493-32488	Sequence 32488, A
c 22	1.7	1487	12	US-10-062-674-1944	Sequence 1944, Ap
c 23	1.7	2577	10	US-09-938-842A-337	Sequence 337, App
c 24	1.7	2577	12	US-09-938-842A-337	Sequence 337, App
c 25	1.7	3086	12	US-10-369-493-37151	Sequence 27151, A
c 26	1.7	3757	12	US-10-262-445-59	Sequence 59, Appl
c 27	1.7	4721	12	US-10-262-445-57	Sequence 57, Appl
c 28	1.7	14353	13	US-10-311-455-3408	Sequence 2408, Ap
c 29	1.7	162450	13	US-10-126-704-1	Sequence 1, Appli
c 30	1.7	162450	15	US-10-071-179-1	Sequence 1, Appli
c 31	1.7	344	9	US-09-770-791-787	Sequence 787, App
c 32	1.7	385	10	US-09-960-352-5500	Sequence 5500, Ap
c 33	1.7	396	10	US-09-938-842A-370	Sequence 370, App
c 34	1.7	396	12	US-09-938-842A-370	Sequence 370, App
c 35	1.7	402	10	US-09-960-352-1265	Sequence 1265, App
c 36	1.7	406	10	US-09-960-352-273	Sequence 273, App
c 37	1.7	435	11	US-09-918-995-1982	Sequence 1982, Ap
c 38	1.7	442	15	US-10-178-213-328	Sequence 328, App
c 39	1.7	490	11	US-09-770-961-675	Sequence 675, App
c 40	1.7	498	9	US-09-244-694-116	Sequence 116, App
c 41	1.7	543	13	US-10-027-632-323443	Sequence 323443,
c 42	1.7	543	13	US-10-027-632-323443	Sequence 323443,
c 43	1.7	543	14	US-10-027-632-323443	Sequence 323443,
c 44	1.7	543	14	US-10-027-632-323443	Sequence 323443,
c 45	1.7	563	10	US-09-796-692-6933	Sequence 6933, Ap

ALIGNMENTS

RESULT 1

US-09-938-842A-1034
Patent No. US20030160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1034
LENGTH: 1071
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1034

Query Match	100.0%	Score 1071,	DB 10,	Length 1071,
Best Local Similarity	100.0%	Pred No. 0,		
Matches 1071,	Conservative 0,	Mismatches 0,	Indels 0,	Gaps 0,
QY	1	ATGGCGACAATTCAGAGCTTGAAGAAGTTTCAGCGCAAGATCAAACTCTAAGACGCGTT	60	
DB	1	ATGGCGACAATTCAGAGCTTGAAGAAGTTTCAGCGCAAGATCAAACTCTAAGACGCGTT	60	
QY	61	GATCTAACCATCATCAACGGCGTTCAGAAAGTTCAGAACTTCAGACCTTCCAGTAAAT	120	
DB	61	GATCTAACCATCATCAACGGCGTTCAGAAAGTTCAGAACTTCAGACCTTCCAGTAAAT	120	

721 ATCGTGGTCCGTCGATCAGCTCAGTTATAGCTTTTCCCGCGCGCTGCTTCGCG 780
781 TCGTCTTACGTCGCGCTGTTTCAACAGAGCTTCCAGATGGGTAGACACCTCCCTTTACAA 840
781 TCGTCTTACGTCGCGCTGTTTCAACAGAGCTTCCAGATGGGTAGACACCTCCCTTTACAA 840
841 GTTGTTCGAAGCAGCGCTTTGTATCCGTTTTCAGAGCTTACGGTTCGAATTTATCAAGA 900
841 GTTGTTCGAAGCAGCGCTTTGTATCCGTTTTCAGAGCTTACGGTTCGAATTTATCAAGA 900
901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGGCGTAAACAACCGGTAGTTTCATCGTCA 960
901 GCGAGCTCGGTTATGGCTCCGAGCTCAGGCTCAGGCGTAAACAACCGGTAGTTTCATCGTCA 960
961 ATTGCAACAACAACGACGACACCGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACA 1020
961 ATTGCAACAACAACGACGACACCGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACA 1020
1021 GAGCTTCCAGCTTATGAGCACCACACGAGCTCAGGCTCATCGAACCCACTGA 1071
1021 GAGCTTCCAGCTTATGAGCACCACACGAGCTCAGGCTCATCGAACCCACTGA 1071

RESULT 3

S-09-924-035A-502/c

Sequence 502, Application US/09924035A

Patent No. US20020142319A1

GENERAL INFORMATION:

APPLICANT: Grilach, Jörn

TITLE OF INVENTION: Expressed Sequences of Arabidopsis

FILE OF INVENTION: thaliana

CURRENT APPLICATION NUMBER: US/09/924,035A

CURRENT FILING DATE: 2000-08-11

PRIOR APPLICATION NUMBER: US 60/148,784

PRIOR FILING DATE: 1999-08-13

NUMBER OF SEQ ID NOS: 900

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 502

LENGTH: 460

TYPE: DNA

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)-(460)

OTHER INFORMATION: n = A,T,C or G

S-09-924-035A-502

Query Match 27.0%; Score 289; DB 10; Length 460;
Best Local Similarity 99.6%; Pred. No. 9.1e-147;
Matches 459; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
600 GCAAGCTTGCATCATCCATCTGGTGGCTCAGCACTTCTGCGGCAAGGAATGTATCCGAT 659
460 GCAAGCTTGCATCATCCATCTGGTGGCTCAGCACTTCTGCGGCAAGGAATGTATCCGAT 401
660 GTGGCTATTCCATCAACGCAATGATTCCAGCGTCCGAGCTTCTTCTTGTATCCACA 719
400 GTGGCTATTCCATCAACGCAATGATTCCAGCGTCCGAGCTTCTTCTTGTATCCACA 341
720 AATCGTGGTCCGTCGAATCAGCTCAGTTATAGCTTTTCCCGCCGCGCTGCTTCGCC 779
340 AATCGTGGTCCGTCGAATCAGCTCAGTTATAGCTTTTCCCGCCGCGCGNG-TTCGCC 282
780 GTGCTTTACGTCGCGCTGTTCAACAGGCTTCCAGATGGCTAGACACCTCCTTTACA 839
281 GTGCTTTACGTCGCGCTGTTCAACAGGCTTCCAGATGGCTAGACACCTCCTTTACA 222
840 AGTTGTTCCAGCAGCGCTTTGTATCCGTTTTCAGAGCTTTCGAGATTTATCAAG 899
221 AGTTGTTCCAGCAGCGCTTTGTATCCGTTTTCAGAGCTTTCGAGATTTATCAAG 162

900 AGGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGGCTAAACACCGGTAGTTTCATCGTC 959
Db 161 AGGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGGCTAAACACCGGTAGTTTCATCGTC 102
960 AATTGCAACAACAACGAGCAGCAGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACA 1019
Db 101 AATTGCAACAACAACGAGCAGCAGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACA 42
900 AGAGCTTCCAGCTTCAATGAGCACCACCAACGACCGGTCAAT 1060
Db 41 AGAGCTTCCAGCTTCAATGAGCACCACCAACGACCGGTCAAT 1

RESULT 4

US-09-770-444-615/c

Sequence 615, Application US/09770444

Patent No. US20020023280A1

GENERAL INFORMATION:

APPLICANT: Gorlach, Jörn

APPLICANT: An, Yong-Qiang

APPLICANT: Hamilton, Carol M.

APPLICANT: Price, Jennifer L.

APPLICANT: Raines, Tracy M.

APPLICANT: Yu, Yang

APPLICANT: Rameaka, Joshua G.

APPLICANT: Page, Amy

APPLICANT: Matthew, Abraham V.

APPLICANT: Ledford, Brooke L.

APPLICANT: Woessner, Jeffrey P.

APPLICANT: Haas, William David

APPLICANT: Garcia, Carlos A.

APPLICANT: Kriker, Maja

APPLICANT: Slader, Ted

APPLICANT: Davis, Keith R.

APPLICANT: Allen, Keith

APPLICANT: Hoffman, Neil

APPLICANT: Hurlan, Patrick

TITLE OF INVENTION: Expressed Sequences of Arabidopsis

FILE OF INVENTION: thaliana

FILE REFERENCE: 2027 (PARA-016PRV)

CURRENT APPLICATION NUMBER: US/09/770,444

CURRENT FILING DATE: 2001-01-26

PRIOR APPLICATION NUMBER: 60/178,502

PRIOR FILING DATE: 2000-01-27

NUMBER OF SEQ ID NOS: 999

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 615

LENGTH: 453

TYPE: DNA

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)....(453)

OTHER INFORMATION: n = A,T,C or G

US-09-770-444-615

Query Match 26.7%; Score 286; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 3.9e-145;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
775 TCGCGTGGTCTTACGTCGCGCTGTTCAACAGGCTTCCAGATGGGTAGACACCTCCT 834
Db 286 TCGCGTGGTCTTACGTCGCGCTGTTCAACAGGCTTCCAGATGGGTAGACACCTCCT 227
935 TTACAAGTTGTTCCAGCAGCGGCTTTGTATCCGTTTCAGACGTTAGCGGTTCAATT 894
Db 226 TTACAAGTTGTTCCAGCAGCGGCTTTGTATCCGTTTCAGACGTTAGCGGTTCAATT 167
895 TCAAGAGCGAGCTCGGTTATGGTCCGAGCTCAAGCTCAGCGGTAAACACCGGTAGTTCA 954
Db 166 TCAAGAGCGAGCTCGGTTATGGTCCGAGCTCAAGCTCAGCGGTAAACACCGGTAGTTCA 107
955 TCGTCAATTGCAACAACAACGACGCAACGCGTGGAGAGCTTCTCCCTAGAGATATACAG 1014

Db 106 TCGTCAATTGCAACAACACGACGACACGCTGAGAGACTTCTCCCTAGAGATATACGAG 47
QY 1015 AAACAAGAGCTTCACAGTTCATGACGACACACACACGCGTCAT 1060
Db 46 AAACAAGAGCTTCACAGTTCATGACGACACACACACGCGTCAAT 1

RESULT 5

US-09-770-696-257
; Sequence 257, Application US/09770696
; Patent No. US20010044940A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2031US (PARA-020PRV)
; CURRENT APPLICATION NUMBER: US/09/770,696
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,278
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 911
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257
; LENGTH: 185
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-696-257

Query Match 17.3%; Score 185; DB 9; Length 185;

Best Local Similarity 100.0%; Pred. No. 4.9e-90;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 GCAAGATCAAACTTAAGAGCGTTGATCTAACCATCATCAACGGCGTCAGAACTCG 94
Db 1 GCAAGATCAAACTTAAGAGCGTTGATCTAACCATCATCAACGGCGTCAGAACTCG 60
QY 95 AAATTTCAAGACCTTTCCAAAGTAAATCCCAAGTGTCTCGAGCCCAAGCGGACCGG 154
Db 61 AAATTTCAAGACCTTTCCAAAGTAAATCCCAAGTGTCTCGAGCCCAAGCGGACCGG 120
QY 155 TGATGCGGTGCTTTTCAATGCTTTACTCCACCGGTTTCGACGACACACCATTTGAAGA 214
Db 121 TGATGCGGTGCTTTTCAATGCTTTACTCCACCGGTTTCGACGACACACCATTTGAAGA 180
QY 215 GAGCT 219
Db 181 GAGCT 185

RESULT 6

US-10-295-403-147
; Sequence 147, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:

; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omais
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/101,349
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/103,312
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/108,734
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/113,409
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 147
; LENGTH: 1604
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (143)..(1345)
; OTHER INFORMATION: G802
US-10-295-403-147

Query Match 2.1%; Score 23; DB 15; Length 1604;

Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAAGACCGTCACAGAA 242
Db 317 TCGACTAAAGACCGTCACAGAA 339

RESULT 7

US-10-369-493-37704/c
; Sequence 37704, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 37704
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-37704

Query Match

Best Local Similarity 1.9%; Score 20; DB 12; Length 1263;

100.0%; Pred. No. 5.3;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 764 CCGCCGCTCTTCGGCGTCG 783
b 874 CCGCCGCTCTTCGGCGTCG 855

RESULT 8

S-09-963-038A-1
Sequence 1, Application US/09963038A
Publication No. US20030143200A1
GENERAL INFORMATION:
APPLICANT: Tikoo, Suresh K.
TITLE OF INVENTION: PORCINE ADENOVIRUS E1 REGION
FILE REFERENCE: 293102003200
CURRENT APPLICATION NUMBER: US/09/963,038A
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 34094
TYPE: DNA
ORGANISM: Porcine Adenovirus Type 3
S-09-963-038A-1

Query Match 1.8%; Score 19; DB 13; Length 34094;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 850 AGCAGCGGCTTTGTATCCG 868
b 21527 AGCAGCGGCTTTGTATCCG 21545

RESULT 9

S-10-199-550-1
Sequence 1, Application US/10199550
Publication No. US20030099615A1
GENERAL INFORMATION:
APPLICANT: Tikoo, Suresh K.
TITLE OF INVENTION: PORCINE ADENOVIRUS E1 AND E4 REGIONS
FILE REFERENCE: 293102003220
CURRENT APPLICATION NUMBER: US/10/199,550
CURRENT FILING DATE: 2002-12-20
PRIOR APPLICATION NUMBER: US 09/963,038
PRIOR FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 34094
TYPE: DNA
ORGANISM: Porcine Adenovirus Type 3
S-10-199-550-1

Query Match 1.8%; Score 19; DB 15; Length 34094;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 850 AGCAGCGGCTTTGTATCCG 868
b 21527 AGCAGCGGCTTTGTATCCG 21545

RESULT 10

S-10-245-603A-1
Sequence 1, Application US/10245603A
Publication No. US20030130187A1
GENERAL INFORMATION:
APPLICANT: REDDY, Police Seshidhar
APPLICANT: TIKOO, Suresh Kumar
APPLICANT: BABTUK, Lorne A.
TITLE OF INVENTION: PORCINE ADENOVIRUS TYPE 3 GENOME
FILE REFERENCE: 293102002410

; CURRENT APPLICATION NUMBER: US/10/245,603A
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 09/292,034
; PRIOR FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: US 60/081,882
; PRIOR FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 34094
; TYPE: DNA
; ORGANISM: Porcine Adenovirus Type 3
; FEATURE:
; US-10-245-603A-1

Query Match 1.8%; Score 19; DB 16; Length 34094;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 850 AGCAGCGGCTTTGTATCCG 868
Db 21527 AGCAGCGGCTTTGTATCCG 21545

RESULT 11

US-09-294-093B-4716/c
; Sequence 4716, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 4716
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700354854H1
; NAME/KEY: unsure
; LOCATION: 96, 99-100, 198, 205
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-4716

Query Match 1.7%; Score 18; DB 9; Length 277;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 385 GCACGGGTACGGGACG 402
Db 24 GCACGGGTACGGGACG 7

RESULT 12

US-09-918-995-30316/c
; Sequence 30316, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30316
; LENGTH: 364
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(364)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30316

Query Match 1.7%; Score 18; DB 11; Length 364;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 AGCTCTGGCATCATCCAC 620
Db 218 AGCTCTGGCATCATCCAC 201

RESULT 13
US-09-918-995-34549/c
; Sequence 34549, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34549
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-34549

Query Match 1.7%; Score 18; DB 11; Length 377;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 AGCTCTGGCATCATCCAC 620
Db 226 AGCTCTGGCATCATCCAC 209

RESULT 14
US-09-777-564-1604/c
; Sequence 1604, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Marion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 1604
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-777-564-1604

Query Match 1.7%; Score 18; DB 9; Length 424;

Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 AGCTCTGGCATCATCCAC 620
Db 83 AGCTCTGGCATCATCCAC 66

RESULT 15
US-10-015-219-1604/c
; Sequence 1604, Application US/10015219
; Publication No. US2003008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1604
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-015-219-1604

Query Match 1.7%; Score 18; DB 15; Length 424;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 AGCTCTGGCATCATCCAC 620
Db 83 AGCTCTGGCATCATCCAC 66

Search completed: February 2, 2004, 08:33:14
Job time : 808.301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM nucleic - nucleic search, using sw model
Run on: February 2, 2004, 01:11:01 ; Search time 6363.2 Seconds
(without alignments)
10781.591 Million cell updates/sec

Title: US-09-938-842a-3729

Perfect score: 1677

Sequence: 1 gggttaagcgttttactatg.....tttctcgcgtatatattta 1677

Scoring table: OLIGO NUC
Gapop_60.0, Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vt.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vt.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rod.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sv.*
39: em_hgt_hum.*
40: em_hgt_mus.*
41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1677	100.0	1677	6	AX509034
2	1677	100.0	91854	8	AC003680 Arabidops
3	193	11.5	1483	8	AF085354 Arabidops
4	102	6.1	1660	8	AY056214 Arabidops
5	24	1.4	250957	2	AC106200 Rattus no
6	24	1.4	282156	2	AC119331 Rattus no
7	23	1.4	101	6	AX114617 Sequence
8	23	1.4	101	6	AX114618 Sequence
9	23	1.4	101	6	AX114743 Sequence
10	23	1.4	101	6	AX114744 Sequence
11	23	1.4	101	6	AX118672 Sequence
12	23	1.4	101	6	AX118673 Sequence
13	23	1.4	101	6	AX118673 Sequence
14	23	1.4	101	6	AX128089 Sequence
15	23	1.4	101	6	AX128090 Sequence
16	23	1.4	101	6	AX644690 Sequence
17	23	1.4	7240	6	AX114613 Sequence
18	23	1.4	7240	6	AX114739 Sequence
19	23	1.4	7240	6	AX118669 Sequence
20	23	1.4	7240	6	AX128086 Sequence
21	23	1.4	7240	6	AX644687 Sequence
22	23	1.4	7240	9	HUMINSRD
23	23	1.4	7641	9	M29929 Human insul
24	23	1.4	47467	2	AC135456 Rattus no
25	23	1.4	72172	9	AC010311 Homo sapi
26	23	1.4	125227	2	AC102324 Mus muscu
27	23	1.4	140868	2	AC142238 Rattus no
28	23	1.4	174141	2	AC102313 Mus muscu
29	23	1.4	187783	2	AC137462 Rattus no
30	23	1.4	241280	2	AC115158 Rattus no
31	23	1.4	248948	2	AC097421 Rattus no
32	23	1.4	254082	2	AC101655 Mus muscu
33	23	1.4	260809	2	AC126820 Rattus no
34	23	1.4	267375	2	AC095339 Rattus no
35	22	1.3	47328	6	AX059532 Sequence
36	22	1.3	103632	8	AF074021 Arabidops
37	22	1.3	164679	9	AC005921
38	22	1.3	179155	10	AL772401 Mouse DNA
39	22	1.3	189484	2	BX470113 Danio rer
40	22	1.3	200001	8	AL161501 Arabidops
41	22	1.3	216886	2	AC127846 Rattus no
42	22	1.3	236723	2	AC106621 Rattus no
43	21	1.3	330	6	AX210412 Sequence
44	21	1.3	394	6	AX245777 Sequence
45	21	1.3	407	6	AX246703 Sequence

ALIGNMENTS

RESULT 1
AX509034
LOCUS AX509034 1677 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 3729 from Patent WO0216555.
ACCESSION AX509034
VERSION AX509034.1 GI:23390271
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis
1
REFERENCE
AUTHORS Harper J.F., Kreps, J., Wang, X. and Zhu, T.
TITLE Stress-regulated genes of plants, transgenic plants containing

same, and methods of use
Patent: WO 021655-A 3729 28-FEB-2002;
The Scripps Research Institute (US) ; Syngenta Participations AG
(CH)
FEATURES
source Location/Qualifiers
1..1677
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 508 a 292 c 335 g 542 t
ORIGIN
Query Match 100.0%; Score 1677; DB 6; Length 1677;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTAAAGCGTTTACTTATGCTTATATGCAACGGAAGATATGCAATGTTGGAATGC 60
DB 1 GGTAAAGCGTTTACTTATGCTTATATGCAACGGAAGATATGCAATGTTGGAATGC 60
2Y 61 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTTA 120
DB 61 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTTA 120
2Y 121 GAAATGTGTTTATGCAACAGGTAGAGACATAACCATAGACAGATGTCTCGAAGAGA 180
DB 121 GAAATGTGTTTATGCAACAGGTAGAGACATAACCATAGACAGATGTCTCGAAGAGA 180
2Y 181 TAAGCTTCTCTATGTCTAAAGAAATGACCGATACGAATAAACCAAGCATCATTAAGAT 240
DB 181 TAAGCTTCTCTATGTCTAAAGAAATGACCGATACGAATAAACCAAGCATCATTAAGAT 240
2Y 241 TAAATGTTTGTAGAAATACTACATTTATGTAATGTAATGTTGTTAGTGAAGT 300
DB 241 TAAATGTTTGTAGAAATACTACATTTATGTAATGTAATGTTGTTAGTGAAGT 300
2Y 301 AAAAACAATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAATTTATGTTGGCG 360
DB 301 AAAAACAATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAATTTATGTTGGCG 360
2Y 361 TAATGAATGTATGCTGATGTTAGGCAAAAGTTGGTGGTGGCAATTAACAAATATC 420
DB 361 TAATGAATGTATGCTGATGTTAGGCAAAAGTTGGTGGTGGCAATTAACAAATATC 420
QY 421 CCTCTGTGTGGACCCGAACTGTAACTCGGAAAGTGGAAACCACTGTTGTTTAACTTT 480
DB 421 CCTCTGTGTGGACCCGAACTGTAACTCGGAAAGTGGAAACCACTGTTGTTTAACTTT 480
QY 481 AAGCTAAAGGTTTACTACCGGTTTACCGGTTTAAATTTGGTGTGTTTAAATCTAATCCC 540
DB 481 AAGCTAAAGGTTTACTACCGGTTTACCGGTTTAAATTTGGTGTGTTTAAATCTAATCCC 540
QY 541 GGATCCGTTTGTGTTTAACTCAAGCCACGTTATCGCCCAATTTTGTGTTTGGTGG 600
DB 541 GGATCCGTTTGTGTTTAACTCAAGCCACGTTATCGCCCAATTTTGTGTTTGGTGG 600
QY 601 GGTAGGAATGTTGGGTCGAATAGTTGGGCTAGCCCTCAACAAATGTGGAATCGAAG 660
DB 601 GGTAGGAATGTTGGGTCGAATAGTTGGGCTAGCCCTCAACAAATGTGGAATCGAAG 660
QY 661 AGATAGGTTCCAGCTCAGGCCCAATTCACATTTTGGTTTGGTACCTTCTTTTCTGG 720
DB 661 AGATAGGTTCCAGCTCAGGCCCAATTCACATTTTGGTTTGGTACCTTCTTTTCTGG 720
QY 721 TGCTTACGGTCCCTCTTCTGTTGCGGTGCTATGTACAAAGTAGCATAGTGGTTCAAA 780
DB 721 TGCTTACGGTCCCTCTTCTGTTGCGGTGCTATGTACAAAGTAGCATAGTGGTTCAAA 780
QY 781 CCGGAACAAGTACCAACGAATCAAAATAGTTTGAATCGGTTTACATCTAGTTACCGTCG 840
DB 781 CCGGAACAAGTACCAACGAATCAAAATAGTTTGAATCGGTTTACATCTAGTTACCGTCG 840
QY 841 AACTTACAATCATTCGATACITTTGATCTGATTTCTAGTTCGGTTTGTATGTTAATAT 900

DB 841 AACTTACAATCATTCGATACITTTGATCTGATTTCTAGTTGGTTTGTATGTTAATAT 900
QY 901 CCGGATTTGTAACAGTACACAGTACATTAAGTATGCGTATATGATGTCGCGGTTTAATC 960
DB 901 CCGGATTTGTAACAGTACACAGTACATTAAGTATGCGTATATGATGTCGCGGTTTAATC 960
QY 961 AAAGGACGAGCATAGGAGGATTTTGAATCTCTGGAAGAGGATTTATCCATAGACATA 1020
DB 961 AAAGGACGAGCATAGGAGGATTTTGAATCTCTGGAAGAGGATTTATCCATAGACATA 1020
QY 1021 ATTAGCTTTTGGTGGCGAGCCTTGTGACCTACATTAATGGGGTCCACCCCAAGTATG 1080
DB 1021 ATTAGCTTTTGGTGGCGAGCCTTGTGACCTACATTAATGGGGTCCACCCCAAGTATG 1080
QY 1081 GGCCTTACAGCTTTTCCATTAATTAAGTAAATCTTTTTCCTTAAACCAATAAAAAT 1140
DB 1081 GGCCTTACAGCTTTTCCATTAATTAAGTAAATCTTTTTCCTTAAACCAATAAAAAT 1140
QY 1141 ATTGAAAATCTTTCCAAACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTAC 1200
DB 1141 ATTGAAAATCTTTCCAAACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTAC 1200
QY 1201 AAAGCTAGGTATTTTCTTTCGGAGTGTACTAGTAACTAGTAACTAACCAAGATGAGT 1260
DB 1201 AAAGCTAGGTATTTTCTTTCGGAGTGTACTAGTAACTAGTAACTAACCAAGATGAGT 1260
QY 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTTAAATTAAGTATATCTAATAAAT 1320
DB 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTTAAATTAAGTATATCTAATAAAT 1320
QY 1321 AAAAGAAAACATTTTGTGAAAAGAGAAATAAAGTTTACTTGGACCCCATGTACAGATCG 1380
DB 1321 AAAAGAAAACATTTTGTGAAAAGAGAAATAAAGTTTACTTGGACCCCATGTACAGATCG 1380
QY 1381 TCCCAATTAATATCTGATAGAGATAGACCAATGGAAGTGAATTTTTCACGTGGTACAA 1440
DB 1381 TCCCAATTAATATCTGATAGAGATAGACCAATGGAAGTGAATTTTTCACGTGGTACAA 1440
QY 1441 TCGGAATGTTTCTTAAAGCTCATCGAACATCAGACCGGTTGATTTTCCGCGATCAA 1500
DB 1441 TCGGAATGTTTCTTAAAGCTCATCGAACATCAGACCGGTTGATTTTCCGCGATCAA 1500
QY 1501 AAAGCGTTTGAATATCTTCTCAGTGTCTTCTGCTTCTATATATATATCTCTGACGAGTC 1560
DB 1501 AAAGCGTTTGAATATCTTCTCAGTGTCTTCTGCTTCTATATATATATCTCTGACGAGTC 1560
QY 1561 ACATTTAGTAAATCTCTTGGACCGTGAACCGGTTAAACAGATTTCTTCCATTTGATCC 1620
DB 1561 ACATTTAGTAAATCTCTTGGACCGTGAACCGGTTAAACAGATTTCTTCCATTTGATCC 1620
QY 1621 GCTTTTAAACACTCTGCTGCTCATCTCCACCGTCCGTTTCTCTCAGCTATATTTTA 1677
DB 1621 GCTTTTAAACACTCTGCTGCTCATCTCCACCGTCCGTTTCTCTCAGCTATATTTTA 1677

RESULT 2
AC003680/c
LOCUS
DEFINITION Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence,
complete sequence.
ACCESSION AC003680
VERSION AC003680.3 GI:20197048
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
1 (bases 1 to 91854)
Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Niernman, W.C. and Fraser, C.M.

Arabisopsis thaliana chromosome 2 BAC F17K2 genomic sequence
 Unpublished
 2 (bases 1 to 91854)
 Lin, X.
 Direct Submission
 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 3 (bases 1 to 91854)
 Town, C.D. and Kaul, S.
 Direct Submission
 Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA. cdowen@tigr.org
 On Apr 18, 2002 this sequence version replaced gi:6598396.
 Address all correspondence to: at@tigr.org

BAC clone F17K2 is from Arabidopsis thaliana chromosome 2 and is
 near the molecular marker (s) F1S.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.
 Genes were identified by a combination of several methods: Gene
 prediction programs including Genscan+ (Chris Burge,
<http://CCR-08.mit.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant
 of GlimmerW, see Mihaela Pettea,
<http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html>, and
 GeneSplicer (Mihaela Pettea and Steven Salzberg, contact
 mpettea@tigr.org), searches of the complete sequence against a
 peptide database and the plant EST database at TIGR
 (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to
 indicate the level of evidence for their annotation. Genes with
 similarity to other proteins are named after the database hits.
 Genes without significant peptide similarity but with EST
 similarity are named as unknown proteins. Genes without protein
 or EST similarity, that are predicted by more than two gene
 prediction programs over most of their length are annotated as
 hypothetical proteins. Genes encoding tRNAs are predicted by
 tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 Simple repeats are identified by repeatmasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
 Location/Qualifiers
 1..91854
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="C1C02E07"
 /clone="F17K2"
 complement(1..1281)
 /note="overlap with BAC clone F4L23 (AC002387:1..1281)."
 57..747
 /gene="At2g45480"
 /note="synonym: F17K2.27"
 join(<57..242,335..>747)
 /gene="At2g45480"
 join(57..242,335..747)
 /gene="At2g45480"
 /note="Protein sequence is in conflict with the conceptual
 translation"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAF18607.2"
 /db_xref="GI:20197053"
 /translation="DESKPSVQKFFPEVSDKCLBAKFSNNKNDIARSGWNNMV
 NGGLFHGFHSPDVLQGRGCFRLQGVETNEPGRCTRTGKCKSVLSGQKVC
 KMRGKMKHPVDITNSHENAGFSPLTVAVRVVPCKDGDQKHSVVMGTLPLR
 VSDEKSTSCSTDTITDIALRGDDDEYLSPGV"
 complement(1002..2388)
 /gene="At2g45490"
 /note="synonym: F17K2.2; contains a protein kinase domain
 profile (PDOC00100)"
 complement(join(<1002..1112,1208..1280,1366..1580,
 1655..1750,1886..2128,2260..>2388))

/gene="At2g45490"
 complement(join(1002..1112,1208..1280,1366..1580,
 1655..1750,1886..2128,2260..>2388))
 /gene="At2g45490"
 /codon_start=1
 /product="putative protein kinase"
 /protein_id="AAC06151.1"
 /db_xref="GI:2979542"
 /translation="MGKSTSDAGNTEKQWLSADFEIGRPLGKGFGRVYLAREAKS
 KYVALKVIPEKQIEKYIHHOIRREMEICTSLRHENILRLFGWPHDNERIPLILEYA
 HGSELGVGLKQNGHLEQQQATVIASLSQALAYCHCKCVIHRDIKFNILLHGERLKL
 IAPFGSVSSNNKRTKMGGLDYIAEMVENRHDYAVDNWTGLIILCYLCEFLYGNPFPE
 AESQKOTFRKLIDLSFPFLPNVSEAKNLISQLLVKDPSEKLSLSEKIMQHPWIVKN
 ADPKGVCASTDI"
 complement(3414..5980)
 /gene="At2g45500"
 /note="synonym: F17K2.3; predicted by genscan"
 complement(join(<3414..3616,3639..3825,3916..3965,
 4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,
 5506..5585,5670..>5980))
 /gene="At2g45500"
 complement(join(3414..3616,3639..3825,3916..3965,
 4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,
 5506..5585,5670..5980))
 /gene="At2g45500"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAC06152.1"
 /db_xref="GI:2979543"
 /translation="NSFLRGIIDSFSSILNERSKDPVSSSSSTSESMNGIDGVPVT
 NERIALKGYPLAKEETAKGVRAEWGLHDALLHYNARQINNEATSPSPSYIS
 SEKEKVRYSKINQWQSERLOALGVSENNKRTVAYSSASVSTASRYRKIL
 SQTIVARGVATPRNPDAAPKPVXSGNVYDDKLVEIMINTIVDRSPSVKWDV
 AGLNGAKALLEWILPAKRDLTGLRPARVYLLGLLFPFGNGKTKMLAKAVAS
 EQATPFNVSSSLTSKWDISIMSTRSENEASRLKSEFLIQPDGVTSPDVLVI
 IGATNKPQLDALLRLVRIYVPLPDSNVKLLFKTKLKCQPHSLSDGIDIKVKE
 TEGLYRLCIKKHRTISQVTKRYSQSLQALCEEAAMPFRELGANILITIQANKVLN
 FSVQINVEVCLSLKGLNRCVHVTKISTY"
 6429..8288
 /gene="At2g45510"
 /note="synonym: F17K2.4"
 join(<6429..6950,7026..7310,7386..7724,7806..8006,
 8100..>8288)
 /gene="At2g45510"
 join(6429..6950,7026..7310,7386..7724,7806..8006,
 8100..8288)
 /gene="At2g45510"
 /codon_start=1
 /product="putative cytochrome P450"
 /protein_id="AAC06153.1"
 /db_xref="GI:2979544"
 /translation="NEILTSIAITVATFIVLCPTIYLMIRIFTKSRNDRKRYAPVH
 ATVFLLFHSDELVDYETIAREKRYRFLSPGQSEILTADPRNVHILKTRFONYSK
 GHSSRENADLLGHGIFAVDGEKQRKLSFEFSTRVLDFSCSVFRNASKLVG
 VSEFALSGKAFDAQDLMLRCLDSIFKVGFGVELKCLDGFSGEQGFWEAFDEGNVAT
 SSRFIDPLWKLKWFNIGSKLSIATIDKPVYSLITTKRKLAKONTVVRDIL
 SRFLVESKDPNNNDKYLRIILNFMFTAGDITTAALLSWFLYMLCKNPLVQKIVOE
 IDVTFSHEKTTDNGFVESINEALDEMHYLHAALSETLRLYPVPVDMCAENDV
 LPDHRVSKGNIYYIAFWGRTYITGQDAEFKPERWLKQGLFPQSPSPKLSFHA
 GPRICLGKDFAYRQMKIVSMALLHFFRFMADENSKVIYKRMILTIVDGGHLCAIPR
 TST"
 complement(8412..9632)
 /gene="At2g45520"
 /note="synonym: F17K2.5; supported by full length cDNA:
 Ceres:8220"
 complement(join(<8412..8678,8768..8903,8987..9120,
 9289..9356,9459..>9632))
 /gene="At2g45520"
 complement(join(8553..8678,8768..8903,8987..9120,
 9289..9356,9459..9612))
 /gene="At2g45520"
 /codon_start=1
 /product="expressed protein"


```
/protein_id="AAC06155.1"
/db_xref="GI:2979546"
/translation="MGGKGLKREKNVLAHGGPARLPPPPDRSKODDVPSTLRLMN
YTPSPDSTKOVVEKELKAEVVPVPAATESDSDSVVEKKKREKQWDLREN
ELAIIDRSKRKKYKWEAKQKXKGTEDLRENFKHQIRFGDVQAPLKLIA
VVPKARSTLSASGERLRLQADAIRSKGTARPGVPIFVMWQ
9408..9427
repeat_region
/rpt_family=" (CA)n"
complement(9578..9632)
/rpt_family=" (GA)n"
9822..10798
gene
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"
9822..10798
mRNA
/gene="At2g45530"
9948..10670
CDS
/gene="At2g45530"
/codon_start=1
/product="expressed protein"
/protein_id="AAC06154.1"
/db_xref="GI:2979545"
/translation="MQSSSSSGDDLEKQOQOQKOKPOKOSANESNHLTIIVCN
GDSRELVGQIPPEKEVSLRNGSSHEOCRLQDKEVLRLGQCCRGGLAKHRS
CIDAPFTKGSNOCEICQVAVNVTPEPTQNTYVWRIDSYREERERCGPFLW
VARSILGLGLMLVLSITLGVSLPWNIIIGVIVVGLGTALRLTLEFCIEWSLRA
VHRAVQRAEANNFSNTAIPPAL"
9988..10007
repeat_region
/rpt_family=" (CAG)n"
complement(10833..25573)
/gene="At2g45540"
10007..10007
gene
/notes="synonym: F17K2.7"
complement(join<10833..11272,11438..11560,11636..12378,
12454..12937,13023..13119,13254..13374,13464..13548,
13645..13794,13957..14073,14178..14264,14350..14418,
14520..14576,14730..14813,14926..14995,15396..15538,
15768..16244,16322..16465,16545..17093,17237..17431,
17612..17696,18015..18613,18944..19249,19633..19714,
19972..20048,20151..20264,20423..20533,20796..20930,
100.0%; Score 1677; DB 8; Length 91854;
Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTTAAGCGTTTACTTATGTTTATATGCAACGAGAAATATTGCCANTTTTGAATGC 60
DB 76808 GGTTAAGCGTTTACTTATGTTTATATGCAACGAGAAATATTGCCANTTTTGAATGC 76749
QY 61 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTGA 120
DB 76748 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTGA 76689
QY 121 GAAATTCGTTTATTGCAACAGGTAGACATCAACATAGACAGATGTTCTGAGACA 180
DB 76588 GAAATTCGTTTATTGCAACAGGTAGACATCAACATAGACAGATGTTCTGAGACA 76629
QY 181 TAAGCTTCTCTATGCTTAAAGAAATGACCGATACGAATAAACAAGCATCAATTAAGAT 240
DB 76628 TAAGCTTCTCTATGCTTAAAGAAATGACCGATACGAATAAACAAGCATCAATTAAGAT 76569
QY 241 TAAATGTTTGTAGAAATACATCACTTATTATGTAATTTGTTGGTTAGTGAAGT 300
DB 76568 TAAATGTTTGTAGAAATACATCACTTATTATGTAATTTGTTGGTTAGTGAAGT 76509
QY 301 AAAAATCATCGAATCCAAACCTCAAAATTTACCAATCAGGCCAAATTTATTGATCGTGGCG 360
DB 76508 AAAAATCATCGAATCCAAACCTCAAAATTTACCAATCAGGCCAAATTTATTGATCGTGGCG 76449
QY 361 TAATGAATGTTATGCTGATGGTAGGCAAAAGTTGGTGGTGGCAAAATTAACATATTC 420
DB 76448 TAATGAATGTTATGCTGATGGTAGGCAAAAGTTGGTGGTGGCAAAATTAACATATTC 76369
QY 421 CCTCTGTGTGGACCGCAATCTGTAAATCGGAAGAGTGGNACCCACTTGGTTTAACTTTT 480
```

```
DB 76388 CCTCTGTGTGGACCGCAATCTGTAAATCGGAAGGTGGAAACCACTTGGTTAACTTTT 76329
QY 481 AAGCCTAAAAGTTACTACCGGTTTGAACGGTTTATAATTTGGTTTAAATCTTAATCCC 540
DB 76328 AAGCCTAAAAGTTACTACCGGTTTGAACGGTTTATAATTTGGTTTAAATCTTAATCCC 76269
QY 541 GGATCCGTTTGTGTTGTTAATCTCAAGCCACAGTTATCCCAATATTTTGATTTTGAAGTG 600
DB 76268 GGATCCGTTTGTGTTGTTAATCTCAAGCCACAGTTATCCCAATATTTTGATTTTGAAGTG 76209
QY 601 GGTAGGGAATGGTGGGTTGGAATAGTTGGGCTAGCCCTCAACAATGTGTGGAACCTGAAG 660
DB 76208 GGTAGGGAATGGTGGGTTGGAATAGTTGGGCTAGCCCTCAACAATGTGTGGAACCTGAAG 76149
QY 661 AGATAGGAGTCCAGCTCAGGCCACATTTCAATTTTCGTTTGTAGCTTCTTTCTG 720
DB 76148 AGATAGGAGTCCAGCTCAGGCCACATTTCAATTTTCGTTTGTAGCTTCTTTCTG 76089
QY 721 TGCTTACGGTCCCTCTTTCTGGTCCGTCGTATGTACAAGTAGCATAGCTAGTGGTCAAA 780
DB 76088 TGCTTACGGTCCCTCTTTCTGGTCCGTCGTATGTACAAGTAGCATAGCTAGTGGTCAAA 76029
QY 781 CCGAAGCAAGTACCAAGCATCAAAATAGTTTGAATCGTTACATCTAGTTACCGTGC 840
DB 76028 CCGAAGCAAGTACCAAGCATCAAAATAGTTTGAATCGTTACATCTAGTTACCGTGC 75969
QY 841 AACTTACAATCATTTCCGTTACTTTGATCTGATTTCTAGTTCCGTTTGTATGTTAATAT 900
DB 75968 AACTTACAATCATTTCCGTTACTTTGATCTGATTTCTAGTTCCGTTTGTATGTTAATAT 75909
QY 901 CCGATTTGTACAGTACCAAGTACATAGTATCGGTAATGATGTGACCGGTTTAAATC 960
DB 75908 CCGATTTGTACAGTACCAAGTACATAGTATCGGTAATGATGTGACCGGTTTAAATC 75849
QY 961 AAAGGACGAGCATAGGAGATTTTGAATCTCGAAGAGGAGTATTTCATAGACACTA 1020
DB 75848 AAAGGACGAGCATAGGAGATTTTGAATCTCGAAGAGGAGTATTTCATAGACACTA 75789
QY 1021 ATTAGCTTTTGTGGCGAGCTTGTGACCTACATTAATGAGGCTCAACCCCAAGTATG 1080
DB 75788 ATTAGCTTTTGTGGCGAGCTTGTGACCTACATTAATGAGGCTCAACCCCAAGTATG 75729
QY 1081 GCCTTACAGCTTTTTCATATAAATTAAGTAATCTTTTTCCTTAACCAATTAATAAT 1140
DB 75728 GCCTTACAGCTTTTTCATATAAATTAAGTAATCTTTTTCCTTAACCAATTAATAAT 75669
QY 1141 ATTGAAATCTTTCACACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTAC 1200
DB 75668 ATTGAAATCTTTCACACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTAC 75609
QY 1201 AAAGCTAGGTATTTCAATTTGGGAGTGACTAGTAACTAGTAACTAGCAGAAATGCT 1260
DB 75608 AAAGCTAGGTATTTCAATTTGGGAGTGACTAGTAACTAGTAACTAGCAGAAATGCT 75549
QY 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTCTAAACAAT 1320
DB 75548 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTCTAAACAAT 75489
QY 1321 AAAAGAAAAACATTTTGTGAAAAGAGAAATTAAGTTTACTGGAACCCCTGTACAGATGG 1380
DB 75488 AAAAGAAAAACATTTTGTGAAAAGAGAAATTAAGTTTACTGGAACCCCTGTACAGATGG 75429
QY 1381 TCCCATATAATCTGATAGAAGTAGAGCAATGGAAGTATTTGTTCCAGCTGGTGACAA 1440
DB 75428 TCCCATATAATCTGATAGAAGTAGAGCAATGGAAGTATTTGTTCCAGCTGGTGACAA 75369
QY 1441 TCGAAATGGTCTTTTAAAGCTCATCGAAACATCATCGAGCGTTTGAATTTTCCGCACTAA 1500
DB 75368 TCGAAATGGTCTTTTAAAGCTCATCGAAACATCATCGAGCGTTTGAATTTTCCGCACTAA 75309
QY 1501 AAAGCGTTGATATACTATTCATCTGTTTCTGCTCTTATATATATATATATATATATAT 1560
```



```

/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF107-08-P04 (R10678)"
/note="This clone is in a modified pBluescript vector
(FUC-1) as a BamHI/XhoI insert.
ecotype: Columbia"
1..1660
/gene="At2g45680"
1..102
/gene="At2g45680"
103..1173
/gene="At2g45680"
/codon_start=1
/evidence="experimental"
/product="putative PCF2 DNA binding protein"
/protein_id="AA107063.1"
/db_xref="GI:15810351"
/translations="MATIOKLEVAGKQOTLRADLTILINGVNVTSPPFOVNPVTS
LEPKAEVMESEFSLAPRSSTGPPPLKRASTKDRHKVGRRIIMPATCARIFQL
TRELKSGDETIWLENAPFAIATATGTGTPAIAVNGVTKIPTTINADSDMGE
NLMKKKRPNSEYIDISDAVAGSGLAPIATITTIQPPALASSTVAQQLPQGM
YPAIPSNAMIPTVGFAPFLIPQIAGSNOPQLIAFPAAASPSVYAAVQAQSTWARP
PIQVVPSSGQFVSDVSGSNLGRATSWAPSSSSGVTGSSSIATITTTTLRDFSL
1174..1660
/gene="At2g45680"
464 a 400 c 385 g 411 t

3'UTR
BASE COUNT 464 a 400 c 385 g 411 t
ORIGIN
Query Match 6.1%; Score 102; DB 8; Length 1660;
Best Local Similarity 100.0%; Pred. No. 1.4e-44;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1576 CTTGGAGCTGTAACGCCGTTAAACAGTCTTCCCATGCTATCCGCTTTAACACTCT 1635
Db 1 CTTGGAGCTGTAACGCCGTTAAACAGTCTTCCCATGCTATCCGCTTTAACACTCT 60

2Y 1636 CTTGGAGCTGTAACGCCGTTCTCTCAGCTATATTTTA 1677
Db 61 CTTGGAGCTGTAACGCCGTTCTCTCAGCTATATTTTA 102

RESULT 5
AC106200/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-20C16, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC106200
AC106200.4 GI:30579164
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
1 (bases 1 to 250957)
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blenkinsburg,K., Blyth,P., Brown,M.,
Bryant,N., Buay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deranc,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,N., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

```

```

Georgiadis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoque,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowalski,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenz,H., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapus,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackelmech,O., Okwunonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plummer,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smaiz,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrell,P., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseca,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Woodden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,P., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 250957)
Worley,K.C.

Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 250957)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23118058.
The sequence in this assembly is a combination of BAC-based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJUV
Center clone name: CH230-20C16
----- Summary Statistics
Assembly program: Atlas 3.0
Consensus quality: 233323 bases at least Q40
Consensus quality: 235744 bases at least Q30

```

```
Consensus quality: 238092 bases at least Q20
Estimated insert size: 250301; sum-of-ctnigs estimation
Quality coverage: 6x in Q20 bases; sum-of-ctnigs estimation

*****
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 247058: contig of 247058 bp in length
* 247059 247158: gap of unknown length
* 247159 248228: contig of 1070 bp in length
* 248229 248328: gap of unknown length
* 248329 250957: contig of 2629 bp in length.
*
* FEATURES
*   Location/Qualifiers
*     1..250957
*       /organism="Rattus norvegicus"
*       /mol_type="genomic DNA"
*       /db_xref="taxon:10116"
*       /clone="CH230-20C16"
*
*   misc_feature
*     1..1145
*       /note="wgs_end_extension
*       clone_end:Sp6"
*       3949..4624
*       /note="clone_boundary
*       clone_end:Sp6
*       site:EcoRI
*       end_sequence:BH318584"
*       complement(243703..244577)
*       /note="clone_boundary
*       clone_end:T7
*       site:EcoRI
*       end_sequence:BH318581"
*       245116..247058
*       /note="wgs_end_extension
*       clone_end:T7"
*
*   misc_feature
*     66339 a 49081 c 49937 g 74548 t 11052 others
*
* QUERY MATCH
*   Query Match 1.4%; Score 24; DB 2; Length 250957;
*   Best Local Similarity 100.0%; Pred.No.0.3;
*   Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* Y 1097 CATAAATTAAGTAAATCTTTT 1120
* b 59625 CATAAATTAAGTAAATCTTTT 59602
*
* RESULT 6
* C119331
* OCUS
* REFINITION
*   Rattus norvegicus clone CH230-291E4, *** SEQUENCING IN PROGRESS
*   **, 5 unordered pieces.
*
* CESSION
* AC119331
* ERSTON
* HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
* EYWORDS
* Rattus norvegicus (Norway rat)
* ORGANISM
* Rattus norvegicus
*   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
*   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
*   Rattus.
*
* REFERENCE
* 1 (bases 1 to 282156)
* Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
* Allen, C., Allien, H., Albrooks, S., Amin, A., Anguiano, D.,
* Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
* Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
* Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
```

```
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Dalgado O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebraegorgis, E., Geer, K., Gill, R., Grady, M., Guerta, W., Guevara, M.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hernandez, K.,
Harvey, Y., Hawlak, P., Hawes, A., Henderson, N., Hernandez, U.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensheva, L., Loulsbeg, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
hwaackeleme, O., Okwuonu, G., Olampunsaogon, A., Pal, S., Parks, K.,
Fasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Flopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Fuazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Shen, H.,
Sanders, N., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, O., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, B., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmari, K.,
Vallas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 282156)
Worley, K.C.
Direct Submission
Submitted (26-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 282156)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23911491.
The sequence in this assembly is a combination of HAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
ctnigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only ctnigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
```

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GUVU
Center clone name: CH230-291E4
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 253046 bases at least Q40
Consensus quality: 256512 bases at least Q30
Consensus quality: 259010 bases at least Q20
Estimated insert size: 261450; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 58359: contig of 58359 bp in length
* 58360 58459: gap of unknown length
* 58460 277096: contig of 218637 bp in length
* 277097 277196: gap of unknown length
* 277197 278294: contig of 1098 bp in length
* 278295 278394: gap of unknown length
* 278395 279912: contig of 1518 bp in length
* 279913 280012: gap of unknown length
* 280013 282156: contig of 2144 bp in length.
* Location/Qualifiers
* 1. 282156
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-291E4"
* 5162 .6036
* /note="clone boundary
* clone_end:17
* site:
* end sequence:BZ166106"
* 58460 .59573
* /note="wgs contig"
* complement(209107..209982)
* /note="clone boundary
* clone_end:Sp6
* site:
* end sequence:BZ166107"
* 275379 .277096
* /note="wgs end extension
* clone_end:Sp6"

BASE COUNT 86729 a 49533 c 48244 g 75698 t 21962 others
ORIGIN

Query Match 1.4%; Score 24; DB 2; Length 282156;
Best Local Similarity 100.0%; Pred.No. 0.3;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1097 CATAAATTAAAGTAATCTTTT 1120
DB 11519 CATAAATTAAAGTAATCTTTT 11542

RESULT 7
AX114617/c AX114617 101 bp DNA linear PAT 11-MAY-2001
LOCUS
DEFINITION Sequence 29 from Patent WO0128540.
ACCESSION AX114617
VERSION AX114617.1 GI:14031560

KEYWORDS
SOURCE

Homo sapiens (human)
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Purvis, I.J. and McCarthy, L.C.
Therapy of cephalic pain
Patent: WO 0128540-A 29-26-APR-2001;
GLAXO GROUP LIMITED (GB)
Location/Qualifiers
1. 101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source

BASE COUNT 26 a 15 c 14 g 46 t
ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred.No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCAATTAAGATTAAAT 245
DB 75 ACAAGCATCAATTAAGATTAAAT 53

RESULT 8
AX114618/c

LOCUS
DEFINITION Sequence 30 from Patent WO0128540.
ACCESSION AX114618
VERSION AX114618.1 GI:14031561
KEYWORDS
SOURCE
Homo sapiens (human)
Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Purvis, I.J. and McCarthy, L.C.
Therapy of cephalic pain
Patent: WO 0128540-A 30-26-APR-2001;
GLAXO GROUP LIMITED (GB)
Location/Qualifiers
1. 101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source

BASE COUNT 27 a 14 c 14 g 46 t
ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred.No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCAATTAAGATTAAAT 245
DB 75 ACAAGCATCAATTAAGATTAAAT 53

RESULT 9
AX114743/c

LOCUS
DEFINITION Sequence 29 from Patent WO0128539.
ACCESSION AX114743
VERSION AX114743.1 GI:14031685
KEYWORDS
SOURCE
Homo sapiens (human)
Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Purvis, I.J. and McCarthy, L.C.

TITLE Agent for treating cephalic pain
JOURNAL Patent: WO 0128539-A 29 26-APR-2001;
GLAXO GROUP LIMITED (GB)

FEATURES

source
Location/Qualifiers
1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
26 a 15 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 223 ACAAGCATCATTAAAGATTAAAT 245
|||||
Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 10

AX114744/c
LOCUS AX114744 101 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 30 from Patent WO0128539.
ACCESSION AX114744
VERSION AX114744.1 GI:14031686
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Purvis, I. J. and McCarthy, L. C.
TITLE Agent for treating cephalic pain
JOURNAL Patent: WO 0128539-A 30 26-APR-2001;
GLAXO GROUP LIMITED (GB)

FEATURES

source
Location/Qualifiers
1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
27 a 14 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 223 ACAAGCATCATTAAAGATTAAAT 245
|||||
Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 11

AX118672/c
LOCUS AX118672 101 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 28 from Patent WO0129255.
ACCESSION AX118672
VERSION AX118672.1 GI:14035623
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Purvis, I. J. and McCarthy, L. C.
TITLE Diagnostic test for cephalic pain
JOURNAL Patent: WO 0129255-A 28 26-APR-2001;
GLAXO GROUP LIMITED (GB)

FEATURES

source
Location/Qualifiers
1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"

BASE COUNT 26 a 15 c 14 g 46 t
ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245
|||||
Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 12

AX118673/c
LOCUS AX118673 101 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 29 from Patent WO0129255.
ACCESSION AX118673
VERSION AX118673.1 GI:14035624
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Purvis, I. J. and McCarthy, L. C.
TITLE Diagnostic test for cephalic pain
JOURNAL Patent: WO 0129255-A 29 26-APR-2001;
GLAXO GROUP LIMITED (GB)

FEATURES

source
Location/Qualifiers
1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
27 a 14 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245
|||||
Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 13

AX128089/c
LOCUS AX128089 101 bp DNA linear PAT 15-MAY-2001
DEFINITION Sequence 28 from Patent WO0129256.
ACCESSION AX128089
VERSION AX128089.1 GI:14134629
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Purvis, I. J. and McCarthy, L. C.
TITLE Cephalic pain susceptibility marker
JOURNAL Patent: WO 0129256-A 28 26-APR-2001;
GLAXO GROUP LIMITED (GB)

FEATURES

source
Location/Qualifiers
1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
26 a 15 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: February 2, 2004, 07:55:32
Job time : 6370.7 secs

```
Qy 223 ACAAGCATCATTAAAGATTAAAT 245
Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 14
AX128090/c
LOCUS AX128090 101 bp DNA linear PAT 15-MAY-2001
DEFINITION Sequence 29 from Patent WO0129256.
ACCESSION AX128090
VERSION AX128090.1 GI:14134630
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Purvis, I.J. and McCarthy, L.C.
TITLE Cephalic pain susceptibility marker
JOURNAL Patent: WO 0129256-A 29 26-APR-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
source
1..101
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 27 a 14 c 14 g 46 t
ORIGIN
Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245
Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 15
AX644690/c
LOCUS AX644690 101 bp DNA linear PAT 27-FEB-2003
DEFINITION Sequence 28 from Patent WO0233121.
ACCESSION AX644690
VERSION AX644690.1 GI:28610698
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hosford, D. and Purvis, I.J.
TITLE Test for the diagnosis of diabetes and compounds for the treatment
JOURNAL Patent: WO 0233121-A 28 25-APR-2002;
GLAXO GROUP LIMITED (GB)
FEATURES
source
1..101
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 26 a 15 c 14 g 46 t
ORIGIN
Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245
Db 75 ACAAGCATCATTAAAGATTAAAT 53
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

February 1, 2004, 21:59:06 ; Search time 491.261 Seconds
(without alignments)
9214.976 Million cell updates/

itle: US-09-938-842A-3729

effect score:

sequence: 1 ggtaagcgtttacttatg.....tttctctcagctatatattta 1677

coring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

earched: 2552756 seqs, 1349719017 residues

ord size : 0

total number of hits satisfying chosen parameters: 5105512

```

inimum DB sec length: 0

```

aximum	DB sec	length: 0
aximum	DB sec	length: 2000000000

ost-processing: Listing first 45 summaries

atabase : N Genesq 19Jun03: *

- 1: /SIDS1/gcgdata/genseq/genseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/genseq/genseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/genseq/genseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/genseq/genseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/genseq/genseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/genseq/genseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/genseq/genseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/genseq/genseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/genseq/genseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/genseq/genseq-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/genseq/genseq-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/genseq/genseq-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/genseq/genseq-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/genseq/genseq-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/genseq/genseq-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/genseq/genseq-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/genseq/genseq-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/genseq/genseq-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/genseq/genseq-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/genseq/genseq-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/genseq/genseq-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/genseq/genseq-emb1/NA2000A.DAT.*
- 23: /SIDS1/gcgdata/genseq/genseq-emb1/NA2000B.DAT.*
- 24: /SIDS1/gcgdata/genseq/genseq-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/genseq/genseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result	No.	*			DB	ID	Description
		Score	Match	Length			
	1	1677	100.0	1677	24	ABZ15924	Arabidopsis thalia
	2	193	11.5	1483	21	AAH58782	Arabidopsis thalia
C	3	23	1.4	101	22	AAH50475	Insulin receptor m
C	4	23	1.4	101	22	AAH31172	Human insulin rece
C	5	23	1.4	101	22	AAH31269	Human insulin rece
C	6	23	1.4	101	22	AAH04594	INSB polymorphism
C	7	23	1.4	101	22	ABQ72727	Human insulin rece
C	8	23	1.4	7240	22	AAH50570	Insulin receptor g

C	9	23	1.4	7240	22	AH31170	Human insulin rece
C	10	23	1.4	7240	22	AH31267	Human insulin rece
C	11	23	1.4	7240	22	AD04467	Human insulin rece
C	12	23	1.4	7240	24	ABQ72725	Human insulin rece
C	13	22	1.3	17646	25	ABZ74645	Secreted protein g
C	14	22	1.3	17646	25	ABT17023	Human secreted pro
C	15	22	1.3	17646	25	ABZ68163	Human secreted pro
C	16	22	1.3	1082138	21	AAF22305	Arabidopsis thalia
C	17	21	1.3	330	22	AH81545	Human differential
C	18	21	1.3	330	22	AAS37649	Novel human diagno
C	19	21	1.3	407	22	AAS38575	Novel human diagno
C	20	21	1.3	1332	22	AH17158	Human cDNA sequenc
C	21	1.3	2000	22	AH81797	Human differential	
C	22	21	1.3	2305	22	AH15783	Human cDNA sequenc
C	23	21	1.3	2531	22	AS13793	Human cDNA encodin
C	24	21	1.3	2635	23	AAS82940	Human cDNA encodin
C	25	21	1.3	2704	21	AACT6400	DNA encoding novel
C	26	20	1.2	768	23	AAS56350	Human ORF ORF1955
C	27	19	1.1	19	21	AA88786	Salmonella typhi D
C	28	19	1.1	19	21	AA88786	Arabidopsis thalia
C	29	19	1.1	377	25	ABX5749	Arabidopsis thalia
C	30	19	1.1	400	18	AAV78209	Bovine EST associa
C	31	19	1.1	406	21	AAH30796	Staphylococcus aur
C	32	19	1.1	408	22	AH28817	Human colon cancer
C	33	19	1.1	576	22	AH10030	Drosophila melanog
C	34	19	1.1	700	22	AH92907	Human cDNA clone {
C	35	19	1.1	903	23	AAS50398	Human inflammatory
C	36	19	1.1	909	23	AA51756	Staphylococcus aur
C	37	19	1.1	921	23	AA554565	Staphylococcus aur
C	38	19	1.1	1630	25	ABZ09980	Haematopoietic cel
C	39	19	1.1	2000	24	ABZ16028	Arabidopsis thalia
C	40	19	1.1	2352	24	ABZ13787	Arabidopsis thalia
C	41	19	1.1	2397	24	ABZ14380	Arabidopsis thalia
C	42	19	1.1	2500	23	ABU16703	Drosophila melanog
C	43	19	1.1	2925	23	ABU08470	Drosophila melanog
C	44	19	1.1	5011	24	ABU15030	Drosophila melanog
C	45	19	1.1	5901	24	ABU70273	Chemically treated

ALIGNMENTS

RESULT 1

RESUL I
ABZ15924

ID ABZ15924 standard; DNA; 1677 BP.

XX
XX

AC ABZ15924;

22 XX

DT 21-JAN-20

DE Arabidops

XX

KW Arabidops

XX

OS Arabidops
vv

XX
DN

PN
YY
WOZ002166

XX
PD 28-FEB-20XX
07-877-8724-AUG-20
PFXX
XX
07-POW-47 JF

PR 24-AUG-20

PR 26-JAN-20

22-JUN-20
PR

PA (Scri) S

T 19-FEB-2001 (first entry)
X Arabidopsis thaliana SSEE1 cDNA.
X SSEE1; shrunken seed gene; storage reserve; storage protein;
W oil body; transgenic plant; ss.
X Arabidopsis thaliana.
S
X
H Key Location/Qualifiers
H 122..1225
T CDS /*tag= a
T
X WO200061735-A1.
N
X 19-OCT-2000.
D
X 07-APR-2000; 2000WO-US09192.
F
X 08-APR-1999; 99US-0128651.
R
X (GHEO) GEN HOSPITAL CORP.
A
X Lin Y;
I
X WPI: 2000-679483/66.
R P-PSDB; AAB19718.
R
X Novel shrunken seed gene useful for producing transgenic plants having
T altered production of food storage reserve material, intracellular
T transport of storage protein and formation of protein or oil bodies -
S Claim 8; Page 57; 64pp; English.
X
X The present sequence is that of Arabidopsis thaliana SSEE1 (shrunken
C seed) cDNA, which was isolated from a seedling cDNA library using
C a partial genomic clone as probe. SSEE1 encodes a protein (see
C AAB19718) that, when expressed in a cell of a plant, modifies or
C alters the production of a food storage reserve material (e.g.
C protein, lipid or carbohydrate storage reserve), facilitates the
C intracellular transport of a storage protein, or facilitates the
C formation of protein or oil bodies. The invention provides a
C transgenic plant (or plant cell, plant tissue, plant organ or
C plant component) which includes a recombinant SSEE1 transgene that
C modifies the production of food storage reserves, thereby
C increasing nutritional value. An antisense construct is useful for
C modifying destockation tolerance.
X
X Sequence 1483 BP; 461 A; 268 C; 347 G; 407 T; 0 other;
C
X Query Match 11.5%; Score 193; DB 21; Length 1483;
C Best Local Similarity 99.6%; Pred No. 5,9e-85;
C Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Y 1 GGTAAAGCGTTTACTTATGGTTTATATGCAACGGAATATTGCCATTGTTGGAATGC 60
b 1236 GGTAAAGCGTTTACTTATGGTTTATATGCAACGGAATATTGCCATTGTTGGAATGC 1285
Y 61 TTTTTCAGATCATCAAGGCTCCTACAGATTCTTAGGGAATGTTTCAGGCTTTTGTGA 120
b 1286 TTTTTCAGATCATCAAGGCTCCTACAGATTCTTAGGGAATGTTTCAGGCTTTTGTGA 1345
Y 121 GAAATGTGTTTATTGCAACAGGTAGAGAAACATACCATAGACAGATGTATCTCGAAGAGA 180
b 1346 GAAATGTGTTTATTGCAACAGGTAGAGAAACATACCATAGACAGATGTATCTCGAAGAGA 1405
Y 181 TAAGCTTCTATGTCCTAAGAAATGGCCGATACGAAATAAACAGCATCATTTAAAGAT 240
b 1406 TAAGCTTCTATGTCCTAAGAAATGGCCGATACGAAATAAACAGCATCATTTAAAGAT 1465
Y 241 TAAA 244
b 1466 TAAA 1469

RESULT 3

AAH50475/c
ID AAH50475 standard; DNA; 101 BP.
XX
AC AAH50475;
XX
XX 21-AUG-2001 (first entry)
XX Insulin receptor migraine associated polymorphic site #2.
DE Insulin receptor; polymorphic site; single nucleotide polymorphism;
KW SNP; migraine; cephalic pain; insulin receptor signalling pathway;
KW antimigraine; vulnerable; cluster headache; chronic paroxysmal hemiparesis;
KW vascular disorder associated headache; withdrawal; tension headache; ds.
XX
OS Homo sapiens.
XX
XX Location/Qualifiers
FH Key replace(51,A)
FT allele /*tag= a
FT /standard_name= "single nucleotide polymorphism"
XX
XX WO200128539-A2.
PN
XX 26-APR-2001.
PD
XX 19-OCT-2000; 2000WO-GB04031.
PF
XX 19-OCT-1999; 99GB-0024713.
PR
XX 19-OCT-1999; 99US-0160423.
PR
XX (GLAX) GLAXO GROUP LTD.
PA
XX Purvis IJ, McCarthy LC;
XX WPI: 2001-300274/31.
XX
XX Use of agent that modulates directly or indirectly insulin receptor or
PT insulin receptor signaling pathway in the manufacture of medicament for
PT preventing or treating cephalic pain -
XX
XX Example; Page 34; 58pp; English.
XX
CC The present invention describes the use of an agent (I) that modulates
CC directly or indirectly the insulin receptor or insulin receptor
CC signalling pathway in the manufacture of a medicament for preventing or
CC treating cephalic pain. Also described is an isolated polynucleotide (II)
CC or protein (III) comprising a polymorphism that causes susceptibility to
CC cephalic pain, or a naturally occurring polymorphism that is in linkage
CC disequilibrium with the first polymorphism. (I) has antimigraine and
CC vulnerary activities. (I) is useful for treating cephalic pain which may
CC be a cluster headache, chronic paroxysmal hemiparesis, headache
CC associated with vascular disorders, headache associated with substances
CC or their withdrawal (for example drug withdrawal), tension headache and
CC in particular migraine with or without aura. The present sequence
CC represents an insulin receptor oligonucleotide containing a migraine
CC associated polymorphic site, which is used in the exemplification of
CC the present invention.
XX
XX Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;
SQ
Query Match 1.4%; Score 23; DB 22; Length 101;
C Best Local Similarity 100.0%; Pred. No. 0.69;
C Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 223 ACAAGCATCATTTAAAGATTAAAT 245
b 75 ACAAGCATCATTTAAAGATTAAAT 53

RESULT 4

AAH31172/c
ID AAH31172 standard; DNA; 101 BP.

XX AC

XX AAH31172;

XX DT 25-JUL-2001 (first entry)

XX XX

DE Human insulin receptor gene polymorphic site INSBa.

XX XX

KW Human; insulin receptor; diagnosis; cephalic pain; susceptibility;

KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX XX

XX OS

XX Homo sapiens.

XX XX

PH Key

FT variation

FT replace(51,A)

FT /*tag= a

FT /standard_name= "Single nucleotide polymorphism"

XX WO200129255-A2.

XX PD

XX 26-APR-2001.

XX XX

XX 19-OCT-2000; 2000WO-GB04024.

XX XX

XX 19-OCT-1999; 99GB-0024717.

XX XX

XX (GLAX) GLAXO GROUP LTD.

XX PA

XX Purvis IJ, McCarthy LC;

XX PI

XX WPI; 2001-328499/34.

XX DR

XX Diagnosis of susceptibility to migraine in individuals comprises the

XX PT identification of polymorphisms in the insulin receptor gene region

XX XX

XX Example; Page 19; 41pp; English.

XX XX

XX The present sequence is provided in a specification relating to a

XX CC method of diagnosing susceptibility to cephalic pain. Polymorphisms

XX CC in the insulin receptor gene have been found to cause susceptibility to

XX CC cephalic pain, particularly to migraine. The method comprises typing

XX CC a region of the insulin receptor gene or insulin receptor protein of

XX CC an individual. A susceptibility to cephalic pain may be diagnosed

XX CC using a probe, primer or antibody which is capable of detecting a

XX CC polymorphism in an insulin receptor gene region or the insulin

XX CC receptor protein. The present sequence corresponds to a polymorphic

XX CC site and flanking sequences of a single nucleotide polymorphism

XX CC (SNP) in the human insulin receptor gene.

XX XX

SQ Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;

Query Match 1.4%; Score 23; DB 22; Length 101;

Best Local Similarity 100.0%; Pred. No. 0.69;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 223 ACAAGCATCATTAAGATTAAAT 245

Db 75 ACAAGCATCATTAAGATTAAAT 53

RESULT 5

AAH31269/c

ID AAH31269 standard; DNA; 101 BP.

XX AC

XX AAH31269;

XX XX

XX 25-JUL-2001 (first entry)

XX DT

XX Human insulin receptor gene polymorphic site INSBa.

XX DE

XX Human; insulin receptor; antimigraine; cephalic pain; diagnosis;

XX KW susceptibility; single nucleotide polymorphism; SNP; SNP detection;

XX XX

KW migraine; ds.

XX XX

XX OS

XX Homo sapiens.

XX XX

XX Key

FT variation

FT replace(51,A)

FT /*tag= a

FT /standard_name= "Single nucleotide polymorphism"

XX WO200129255-A2.

XX PD

XX 26-APR-2001.

XX XX

XX 19-OCT-2000; 2000WO-GB04050.

XX XX

XX 19-OCT-1999; 99GB-0024717.

XX XX

XX (GLAX) GLAXO GROUP LTD.

XX PA

XX Purvis IJ, McCarthy LC;

XX PI

XX WPI; 2001-316247/33.

XX DR

XX Diagnosing susceptibility to cephalic pain such as migraine by typing

XX PT insulin receptor gene or protein in vivo, or in a biological sample and

XX PT determining individual's susceptibility to cephalic pain

XX XX

XX Example; Page 21; 46pp; English.

XX XX

XX The present sequence is provided in a specification relating to a

XX CC method for diagnosing susceptibility to cephalic pain in an individual.

XX CC The method comprises typing in vivo or in a sample from the

XX CC individual, the insulin receptor gene region or insulin receptor

XX CC protein and thus determining whether the individual is susceptible to

XX CC cephalic pain. Polymorphisms in the insulin receptor gene that

XX CC cause susceptibility to cephalic pain have been found. Susceptibility

XX CC may be diagnosed using a probe, primer or antibody which is capable of

XX CC detecting an insulin receptor gene region or insulin receptor protein

XX CC polymorphism. The method is useful for diagnosing susceptibility to

XX CC migraine, cluster headache, chronic paroxysmal hemianopia, headache

XX CC associated with vascular disorders, headache associated with substances

XX CC or their withdrawal, tension headache, and so on. It is useful for

XX CC assessing the efficacy of agents in relieving cephalic pain, and

XX CC can be used to assess the ability of agents to modulate insulin

XX CC receptor signalling activity. The method may also be used to assess

XX CC the predisposition and/or susceptibility of an individual to the

XX CC development of diseases mediated by the insulin receptor and in the

XX CC development of new drug therapies which selectively target one or more

XX CC allelic variants of the insulin receptor gene. The present sequence

XX CC corresponds to a polymorphic site and flanking sequences of a single

XX CC nucleotide polymorphism (SNP) in the human insulin receptor gene.

XX XX

SQ Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;

Query Match 1.4%; Score 23; DB 22; Length 101;

Best Local Similarity 100.0%; Pred. No. 0.69;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 223 ACAAGCATCATTAAGATTAAAT 245

Db 75 ACAAGCATCATTAAGATTAAAT 53

RESULT 6

AA04594/c

ID AA04594 standard; DNA; 101 BP.

XX AC

XX AA04594;

XX XX

XX 04-JUL-2001 (first entry)

XX DT

XX INSBa polymorphism of human insulin receptor gene.

XX XX

Human; insulin receptor; cephalic pain; therapy; headache;
chronic paroxysmal hemiparesis; vascular disorder; tension headache;
migraine; single nucleotide polymorphism; SNP; ds.

Homo sapiens.

Key Location/Qualifiers
variation replace (SI, A)
/*tag= a

/standard_name= "single nucleotide polymorphism"

WO200128540-A2.

26-APR-2001.

19-OCT-2000; 2000WO-GB04051.

19-OCT-1999; 99GB-0024712.

19-OCT-1999; 99US-0160418.

(GLAX) GLAXO GROUP LTD.

Purvis IJ, McCarthy LC;

WPI; 2001-290815/30.

Use of agent that modulates directly or indirectly insulin receptor or
insulin receptor signaling pathway in manufacture of medicament for
preventing or treating cephalic pain

Claim 7; Page 21; 46pp; English.

The present invention relates to the use of an agent that modulates
directly or indirectly the insulin receptor or insulin receptor
signalling pathway in the manufacture of a medicament for preventing
or treating cephalic pain. Cephalic pain disorders are generally
multifunctional disorder, with a cluster of headache, chronic
paroxysmal hemiparesis, headache associated with vascular disorders,
headache associated with substances or their withdrawal (for example
drug withdrawal), tension headache and in particular migraine with aura
or migraine without aura. The treatment of cephalic pain and migraine
involves manipulation of components of the glucose and lipid metabolism
pathways, in particular by manipulation of the insulin receptor. Insulin
receptor is an important component in the regulation of the glucose and
lipid metabolism pathways. Single nucleotide polymorphisms (SNP) in the
insulin receptor gene contribute to susceptibility to cephalic pain.
The present sequence is INSBA polymorphism of human insulin receptor
gene.

Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;

Query Match 1.4%; Score 23; DB 22; Length 101;

Best Local Similarity 100.0%; Pred. No. 0.69;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

223 ACAAGCATCATTAAGATTAAT 245

75 ACAAGCATCATTAAGATTAAT 53

RESULT 7

EQ72727/c

D ABQ72727 standard; DNA; 101 BP.

ABQ72727;

09-SEP-2002 (first entry)

Human insulin receptor gene polymorphism INSBA.

Human; insulin; receptor; diabetes; antidiabetic; ds;
single nucleotide polymorphism; SNP.

OS Homo sapiens.

XX Key Location/Qualifiers

FT variation replace (SI, A)

FT /*tag= a

FT /standard_name= "single nucleotide polymorphism"

PN WO200233121-A2.

XX 25-APR-2002.

XX 19-OCT-2001; 2001WO-GB04660.

XX 19-OCT-2000; 2000GB-0025678.

XX (GLAX) GLAXO GROUP LTD.

XX Hosford D, Purvis IJ;

XX WPI; 2002-500014/53.

XX Diagnosing diabetes or susceptibility to diabetes in individual by
PT typing insulin receptor gene region or insulin receptor protein in
PT sample obtained from individual

PS Claim 5; Page 31; 61pp; English.

XX he invention relates to a novel method for diagnosing diabetes or
CC susceptibility to diabetes in an individual. The method of the invention
CC has antidiabetic activity. The method is useful for diagnosing diabetes
CC or susceptibility to diabetes in an individual. Other methods of the
CC invention are useful for treating diabetes and for treating and
CC preventing diabetes. The present sequence contains a single nucleotide
CC polymorphism (SNP) present in the human insulin receptor gene.

XX Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;

Query Match 1.4%; Score 23; DB 24; Length 101;

Best Local Similarity 100.0%; Pred. No. 0.69;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 ACAAGCATCATTAAGATTAAT 245

DB 75 ACAAGCATCATTAAGATTAAT 53

RESULT 8

AAH50570/c

ID AAH50570 standard; DNA; 7240 BP.

XX AAH50570;

XX AC

DT 21-AUG-2001 (first entry)

DE Insulin receptor gene exons 14-17 including introns SEQ ID NO:25.

XX Insulin receptor; polymorphic site; single nucleotide polymorphism;
KW SNP; migraine; cephalic pain; insulin receptor signalling pathway;
KW antihypertensive; vulnery; cluster headache; chronic paroxysmal hemiparesis;
KW vascular disorder associated headache; withdrawal; tension headache; ds.

OS Homo sapiens.

XX WO200128539-A2.

XX 26-APR-2001.

XX 19-OCT-2000; 2000WO-GB04031.

XX 19-OCT-1999; 99GB-0024713.

XX 19-OCT-1999; 99US-0160423.

XX (GLAX) GLAXO GROUP LTD.

development of new drug therapies which selectively target one or more allelic variants of the insulin receptor gene.

Sequence 7240 BP; 1952 A; 1653 C; 1820 G; 1815 T; 0 other;

Query Match 1.4%; Score 23; DB 22; Length 7240;

Best Local Similarity 100.0%; Pred. No. 0.56; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

223 ACAAGCATCATTAAGATTAAAT 245

|||||

5056 ACAAGCATCATTAAGATTAAAT 5034

RESULT 11

ID04467/c

AAD04467 standard; DNA; 7240 BP.

AAD04467;

04-JUL-2001 (first entry)

Human insulin receptor DNA (from exons 14 to 17).

Human; insulin receptor; cephalic pain; therapy; headache;
chronic paroxysmal hemikrania; vascular disorder; tension headache;
migraine; single nucleotide polymorphism; SNP; ds.

Homo sapiens.

WO200128540-A2.

26-APR-2001.

19-OCT-2000; 2000WO-GB04051.

19-OCT-1999; 99GB-0024712.

19-OCT-1999; 99US-0160418.

(GLAX) GLAXO GROUP LTD.

Purvis IJ, McCarthy LC;

WPI; 2001-290815/30.

Use of agent that modulates directly or indirectly insulin receptor or insulin receptor signaling pathway in manufacture of medicament for preventing or treating cephalic pain -

Disclosure; Page 42-45; 46pp; English.

The present invention relates to the use of an agent that modulates directly or indirectly the insulin receptor or insulin receptor signalling pathway in the manufacture of a medicament for preventing or treating cephalic pain. Cephalic pain disorders are generally multifunctional disorder, with a cluster of headache, chronic paroxysmal hemikrania, headache associated with vascular disorders, headache associated with substances or their withdrawal (for example drug withdrawal), tension headache and in particular migraine with aura or migraine without aura. The treatment of cephalic pain and migraine involves manipulation of components of the glucose and lipid metabolism pathways, in particular by manipulation of the insulin receptor. Insulin receptor is an important component in the regulation of the glucose and lipid metabolism pathways. Single nucleotide polymorphisms (SNP) in the insulin receptor gene contribute to susceptibility to cephalic pain. The present sequence is complete sequence from exon 14 to 17 of human insulin receptor gene. SNP in this sequence contributes to susceptibility to cephalic pain.

Sequence 7240 BP; 1952 A; 1653 C; 1820 G; 1815 T; 0 other;

Query Match

Best Local Similarity 1.4%; Score 23; DB 22; Length 7240;

Matches 23; Conservative 100.0%; Pred. No. 0.56;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCATTAAGATTAAAT 245

|||||

Db 5056 ACAAGCATCATTAAGATTAAAT 5034

RESULT 12

ABQ72725/c

ID ABQ72725 standard; DNA; 7240 BP.

XX

AC ABQ72725;

XX

DT 09-SEP-2002 (first entry)

XX

DE Human insulin receptor gene.

XX

KW Human; insulin; receptor; diabetes; antidiabetic; ds.

XX

OS Homo sapiens.

XX

FN WO200233121-A2.

XX

PD 25-APR-2002.

XX

PF 19-OCT-2001; 2001WO-GB04660.

XX

PR 19-OCT-2000; 2000GB-0025678.

XX

PA (GLAX) GLAXO GROUP LTD.

XX

PI Hosford D, Purvis IJ;

XX

DR WPI; 2002-500014/53.

XX

PT Diagnosing diabetes or susceptibility to diabetes in individual by

PT typing insulin receptor gene region or insulin receptor protein in

PT sample obtained from individual -

XX

PS Disclosure; Page 57-61; 61pp; English.

XX

CC he invention relates to a novel method for diagnosing diabetes or
CC susceptibility to diabetes in an individual. The method of the invention
CC has antidiabetic activity. The method is useful for diagnosing diabetes
CC or susceptibility to diabetes in an individual. Other methods of the
CC invention are useful for treating diabetes and for treating and
CC preventing diabetes. The present sequence represents the human
CC insulin receptor gene.

XX

SQ Sequence 7240 BP; 1952 A; 1653 C; 1820 G; 1815 T; 0 other;

Query Match

Best Local Similarity 1.4%; Score 23; DB 24; Length 7240;

Matches 23; Conservative 100.0%; Pred. No. 0.56;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCATTAAGATTAAAT 245

|||||

Db 5056 ACAAGCATCATTAAGATTAAAT 5034

RESULT 13

ABZ74645/c

ID ABZ74645 standard; DNA; 17646 BP.

XX

AC ABZ74645;

XX

DT 12-MAY-2003 (first entry)

XX

DE Secreted protein gene 381 genomic fragment HE2CA60, SEQ ID NO:1792.

XX

KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;

KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;

KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;

W drug screening; chromosome identification; chromosome mapping;
 W cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 W antianemic; vulnery; gene; ds.
 S Homo sapiens.

N WO200277013-A2.

D 03-OCT-2002.

F 26-MAR-2002; 2002WO-US09370.

R 27-MAR-2001; 2001US-278650P.

R 12-SEP-2001; 2001US-0950082.

R 12-SEP-2001; 2001US-0950083.

X (HUMA-) HUMAN GENOME SCI INC.

X Rosen CA, Ruben SM;

X WPI; 2003-040578/03.

R New human secreted proteins and nucleic acids, useful for detecting or
 T treating cancer or other hyperproliferative disorders, autoimmune
 T disorders, inflammatory disorders, HIV disease, hepatitis or anemia -
 T Disclosure; Page 2382-2387; 2474pp; English.

X ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 C protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 C ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
 C invention also encompasses antibodies specific for the secreted proteins,
 C the use of the secreted proteins in drug screening and recombinant
 C vectors and host cells comprising a nucleic acid of the invention. The
 C secreted proteins are thought to be involved in biological activities
 C associated with cellular signalling, cellular differentiation, cell
 C migration, prothrombin activation and neurotransmitter activity. The
 C secreted proteins, nucleic acids encoding them, antibodies or antibody
 C fragments specific for the secreted proteins, and modulators of protein
 C activity are useful for diagnosing or treating cancers or other
 C hyperproliferative disorders. Additionally, the secreted proteins and
 C their nucleic acids may also be used in the treatment of autoimmune
 C disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 C (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 C wound healing. Nucleic acids of the invention may be used for chromosome
 C identification, chromosome mapping, in gene therapy, for identifying
 C individuals from minute biological samples, as hybridisation probes, and
 C as molecular weight markers. The present sequence represents a human
 C secreted protein genomic fragment referred to in the disclosure of the
 C invention.

X Sequence 17646 BP; 5370 A; 3099 C; 3561 G; 5616 T; 0 other;

Query Match 1.3%; Score 22; DB 25; Length 17646;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1133 TAAAAATTATTGAAATCTTTC 1154

b 9447 TAAAAATTATTGAAATCTTTC 9426

RESULT 14

ABT17023/C

D ABT17023 standard; DNA; 17646 BP.

X AC ABT17023;

X 03-APR-2003 (first entry)

X Human secreted protein-related DNA sequence - SEQ ID No 377.

X Human; gene; ds; protein therapy; immediate hypersensitivity disease;

KW allergic disorder; asthmatic disorder; gene therapy; secreted protein;
 KW hay fever; allergic conjunctivitis; allergic rhinitis;
 KW binding partner identification; chromosome identification;
 KW radiation hybrid mapping; long-range restriction mapping.

OS Homo sapiens.

XX WO200277188-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002WO-US09239.

XX 27-MAR-2001; 2001US-278650P.

XX 12-SEP-2001; 2001US-0950082.

XX 12-SEP-2001; 2001US-0950083.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-175010/17.

XX Use of human secreted proteins and nucleic acids for preparing a
 PT diagnostic or pharmaceutical composition for diagnosing or treating
 PT allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic
 PT conjunctivitis or rhinitis

PS Disclosure; Page 815-819; 823pp; English.

XX The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for the diagnosis and treatment of allergic disorders, asthmatic
 CC disorders and immediate hypersensitivity diseases (e.g. hay fever,
 CC allergic conjunctivitis and allergic rhinitis). The proteins of the
 CC invention are also useful for identifying a binding partner. The nucleic
 CC acids of the invention are also useful for chromosome identification,
 CC radiation hybrid mapping or long-range restriction mapping. The present
 CC DNA sequence represents a human secreted protein-related DNA sequence.

XX Sequence 17646 BP; 5370 A; 3099 C; 3561 G; 5616 T; 0 other;

Query Match 1.3%; Score 22; DB 25; Length 17646;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 TAAAAATTATTGAAATCTTTC 1154

Db 9447 TAAAAATTATTGAAATCTTTC 9426

RESULT 15

ABZ68163/C

ID ABZ68163 standard; DNA; 17646 BP.

XX AC ABZ68163;

XX 26-MAR-2003 (first entry)

XX Human secreted protein encoding genomic DNA SEQ ID NO 1686.

XX Human; secreted protein; nontropic; neuroprotective; cytostatic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW vulnery; antibacterial; antiparkinsonian; antisking; antianemic;
 KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
 KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
 KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 KW gene therapy; gene; ds.

OS Homo sapiens.

XX WO200277186-A2.

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 2, 2004, 00:53:06 ; Search time 3880.05 Seconds
(without alignments)
10504.671 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 ggttaaggttttactatg.....ttttctcagctatatatta 1677

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_nus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	16.2	317	29	BZ662927
2	245	14.6	267	29	AL761108 Arabidops
3	164	9.8	232	28	BH612074
4	164	9.8	233	28	BH612075

5	150	8.9	278	29	BZ377781
6	149	8.9	347	9	AV521636
7	101	6.0	486	9	AV825375
8	85	5.1	378	10	BE523004
9	82	4.9	146	28	BH617024
10	40	2.4	600	29	CC459772
11	26	1.6	825	28	BH493162
12	22	1.3	207	29	AL764163
13	21	1.3	162	28	AQ055392
14	21	1.3	178	10	BE933474
15	21	1.3	288	10	BF703799
16	21	1.3	289	14	CB477075
17	21	1.3	310	9	AA224031
18	21	1.3	318	9	AW480642
19	21	1.3	329	9	AW358594
20	21	1.3	330	9	AA224071
21	21	1.3	349	14	CD478137
22	21	1.3	365	9	AI466455
23	21	1.3	391	14	CA940873
24	21	1.3	406	10	BF449231
25	21	1.3	413	10	BE207343
26	21	1.3	424	9	AI035951
27	21	1.3	440	10	BF707341
28	21	1.3	440	10	BE255542
29	21	1.3	445	9	AUI84321
30	21	1.3	455	9	AUI84443
31	21	1.3	476	10	BF707342
32	21	1.3	488	13	BX281427
33	21	1.3	521	10	BF704377
34	21	1.3	551	9	AW655154
35	21	1.3	555	10	BF284594
36	21	1.3	564	12	BM384056
37	21	1.3	587	12	BX537127
38	21	1.3	599	10	BG539984
39	21	1.3	635	10	BE252939
40	21	1.3	637	10	BE254600
41	21	1.3	663	28	B53012
42	21	1.3	702	14	CB309574
43	21	1.3	707	13	B0194910
44	21	1.3	711	10	BE251436
45	21	1.3	728	10	BH610568

ALIGNMENTS

RESULT 1

BZ662927

LOCUS

DEFINITION

Arabidopsis thaliana

survey sequence.

ACCESSION

BZ662927

VERSION

BZ662927.1

KEYWORDS

GSS.

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 317)

Alonso, J.M., Leisner, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab

, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.

, Zimmerman, J. and Ecker, J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

BZ662927 317 bp DNA linear GSS 31-JAN-2003
SALK_026421.45.45.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_026421.45.45.x, genomic
survey sequence.

BZ662927 GI:28177691

GSS.

Arabidopsis thaliana

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 317)

Alonso, J.M., Leisner, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab

, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.

, Zimmerman, J. and Ecker, J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of T-DNA. This sequence lies within 300 bases of the 5' end of At2g45680.

Class: T-DNA tagged.

Location/Qualifiers

1. 317
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_024421.45.45.x"
/notes="PCR was performed on Arabidopsis thaliana T-DNA insertion lines each of which contains one or more T-DNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html"

BASE COUNT 86 a 75 c 51 g 105 t

ORIGIN

Query Match 16.2%; Score 271; DB 29; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.8e-126;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1407 GAGCAATGGAAGTATTTGTCACGTGTCACATCGAATGTTCTTTAAAGCTCATCG 1466

2b 27 GAGCAATGGAAGTATTTGTCACGTGTCACATCGAATGTTCTTTAAAGCTCATCG 86

2y 1467 AACACATCAGGACCGTTGATTTTCCCGATCAAAAGCGTTGAATACATTTCTCACTTG 1526

2b 87 AACACATCAGGACCGTTGATTTTCCCGATCAAAAGCGTTGAATACATTTCTCACTTG 146

2y 1527 TTTTCCTGCTCCAT 1586

2b 147 TTTTCCTGCTCCAT 206

2y 1587 AACGCGGTTAAACAGTATCTTCCCATGTCATCGCTTTTACAACTCTCGTCATCT 1646

2b 207 AACGCGGTTAAACAGTATCTTCCCATGTCATCGCTTTTACAACTCTCGTCATCT 266

2y 1647 CCACCGTCCGTTTCTCTCAGCTATATTTTA 1677

2b 267 CCACCGTCCGTTTCTCTCAGCTATATTTTA 297

RESULT 2

AL761108

LOCUS

AL761108.1 Arabidopsis thaliana T-DNA flanking sequence GK-205F06-014511, genomic survey sequence.

ACCESSION

AL761108

VERSION

AL761108.1

KEYWORDS

GSS.

SOURCE

Arabidopsis thaliana

ORGANISM

Arabidopsis thaliana (thale cress)

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE

1

AUTHORS

Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H. and Weisshaar, B.

TITLE

A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines

JOURNAL

Unpublished

REFERENCE

2

AUTHORS

Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.

TITLE

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics

JOURNAL

Unpublished

REFERENCE

3

(bases 1 to 267)

AUTHORS

Direct Submission

JOURNAL

COMMENT

Rosso, M., Li, Y., Strizhov, N. and Weisshaar, B.

Direct Submission

Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone f418. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at:

<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

Location/Qualifiers

1. 267

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-205F06-014511"

/clone_lib="Arabidopsis thaliana T-DNA insertion lines"

/notes="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequences were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 64 a 48 c 68 g 87 t

ORIGIN

Query Match 14.5%; Score 245; DB 29; Length 267;

Best Local Similarity 100.0%; Pred. No. 3.2e-113;

Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 AGTTGGTGGCGGAAATACACATTTATCCCTCTGTGTGACCCGGAATCTGTAATCG 449

Db 23 AGTTGGTGGCGGAAATACACATTTATCCCTCTGTGTGACCCGGAATCTGTAATCG 82

Qy 450 GAAAGGTGGAACCCACTTGGTTTAACTTTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 509

Db 83 GAAAGGTGGAACCCACTTGGTTTAACTTTTAAAGCTTAAAGCTTAAAGCTTAAAGCT 142

Qy 510 GGTTTATAATTTGGTGTGTTTAAATCTTAATCCCGATCCGTTGTTGTTTAACTCTCAAGGCC 569

Db 143 GGTTTATAATTTGGTGTGTTTAAATCTTAATCCCGATCCGTTGTTGTTTAACTCTCAAGGCC 202

Qy 570 ACCTTATCGCAATATTTGATTTTGAAGGAGTGGGATGAGGATGAGGATGAGGATGAGG 629

Db 203 ACCTTATCGCAATATTTGATTTTGAAGGAGTGGGATGAGGATGAGGATGAGGATGAGG 262

Qy 630 GCCTA 634

Db 263 GCCTA 267

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Location/Qualifiers
1..232
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_032102"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 59 a 45 c 46 g 82 t
ORIGIN
Query Match 9.8%; Score 164; DB 28; Length 232;
Best Local Similarity 100.0%; Pred. No. 6.3e-72;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 183 AGCTTCTCTATGCTTAAGAAGATGACCGATACGATTAACAAACAGCATCATTAAGATTA 242
D 188 AGCTTCTCTATGCTTAAGAAGATGACCGATACGATTAACAAACAGCATCATTAAGATTA 129
Y 243 AATGTTTGTGAAGAAATACACTATTATTATGGAATTTGTTGGTTAGTGAAGATTA 302
D 128 AATGTTTGTGAAGAAATACACTATTATTATGGAATTTGTTGGTTAGTGAAGATTA 69
Y 303 AAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAT 346
D 68 AAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAT 25

RESULT 4
H612075/c
OCUS
DEFINITION
B612075 233 bp DNA linear GSS 04-JAN-2002
SALK_032104 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_032104, genomic survey sequence.
CCESION
B612075
EYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 233)
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 453 4100 x1752
Fax: 858 558 6379

Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.
Class: TDNA tagged.

FEATURES
source
Location/Qualifiers
1..233
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_032104"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 58 a 45 c 47 g 83 t
ORIGIN
Query Match 9.8%; Score 164; DB 28; Length 233;
Best Local Similarity 100.0%; Pred. No. 6.3e-72;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 183 AGCTTCTCTATGCTTAAGAAGATGACCGATACGATTAACAAACAGCATCATTAAGATTA 242
Db 188 AGCTTCTCTATGCTTAAGAAGATGACCGATACGATTAACAAACAGCATCATTAAGATTA 129
QY 243 AATGTTTGTGAAGAAATACACTATTATTATGGAATTTGTTGGTTAGTGAAGATTA 302
Db 128 AATGTTTGTGAAGAAATACACTATTATTATGGAATTTGTTGGTTAGTGAAGATTA 69
QY 303 AAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAT 346
Db 68 AAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAT 25

RESULT 5
BZ377781
LOCUS
DEFINITION
BZ377781 278 bp DNA linear GSS 26-NOV-2002
SALK_106185.39.30.x Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_106185.39.30.x, genomic survey sequence.
ACCESSION
BZ377781
VERSION
BZ377781.1 GI:25467878
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 278)
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At2g45680.

FEATURES
source
Location/Qualifiers
1..278
/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK 106185.39.30.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"
72 a 63 c 44 g 93 t
BASE COUNT
ORIGIN

Query Match 8.9%; Score 150; DB 29; Length 278;
Best Local Similarity 100.0%; Pred. No. 8.7e-65;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1528 TTTCCTGCTCTATATATATATCTGACGAGTCACATTAGTAATCTCTTGACGTGA 1587
Db 109 TTTCCTGCTCTATATATATATCTGACGAGTCACATTAGTAATCTCTTGACGTGA 168
QY 1588 AGCCGTTAAACGATCTTTCCCATTTGTATCCGCTTTTAAACAACCTCGTCTCATCTC 1647
Db 169 AGCCGTTAAACGATCTTTCCCATTTGTATCCGCTTTTAAACAACCTCGTCTCATCTC 228
QY 1648 CACGTCGCTTTCTTCACGTATATTTA 1677
Db 229 CACGTCGCTTTCTTCACGTATATTTA 258

RESULT 6
AV521636/c
LOCUS AV521636 347 bp mRNA linear EST 07-SEP-2000
DEFINITION Old Arabidopsis thaliana aboveground organs two to six-week
ACCESSION AV521636
VERSION AV521636.1 GI:8681163
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 347)
AUTHORS Asamizu E., Nakamura Y., Sato S. and Tabata S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
PubMed 10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

Location/Qualifiers
1. 347
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="AP263b12P"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
/note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
102 a 77 c 54 g 114 t
BASE COUNT
ORIGIN

Query Match 8.9%; Score 149; DB 9; Length 347;
Best Local Similarity 100.0%; Pred. No. 2.8e-64;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTAAAGCGTTTACTTATGTTTATGCAAGGAGATATTCGATTGTGGATGC 60
Db 241 GGTAAAGCGTTTACTTATGTTTATGCAAGGAGATATTCGATTGTGGATGC 182
QY 61 TTTTTCAGATCATCAAGGCTCCTACAGATTCTTCTAGGGAATGTTTCAGGCTTTTGTTA 120
Db 181 TTTTTCAGATCATCAAGGCTCCTACAGATTCTTCTAGGGAATGTTTCAGGCTTTTGTTA 122
QY 121 GAAATTGCTTTATTGCAACAGGTAGAGA 149
Db 121 GAAATTGCTTTATTGCAACAGGTAGAGA 93

RESULT 7
AV825375 486 bp mRNA linear EST 01-APR-2002
LOCUS AV825375
DEFINITION AV825375 RAF17 Arabidopsis thaliana cDNA clone RAF107-08-P04 5', mRNA sequence.
ACCESSION AV825375
VERSION AV825375.1 GI:19867435
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 486)
AUTHORS Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: meski@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

Location/Qualifiers
1. 486
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAF107-08-P04"
/dev_stage="rossette plants"
/lab_host="DH10B"
/clone_lib="RAF17"
/note="Site_1: BamHI; Site_2: SalI; subjected to cold-treated (1, 2, 5, 10, 24 hr)."
131 a 118 c 117 g 117 t 3 others
BASE COUNT
ORIGIN

Query Match 6.0%; Score 101; DB 9; Length 486;
Best Local Similarity 100.0%; Pred. No. 8.4e-40;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1577 TTGGACGTGAACCCCGTTAAACGATTTCTTCCCATTTGATCCGTTTAAACATCTC 1636
Db 1 TTGGACGTGAACCGCGTTAAACGATTTCTTCCCATTTGATCCGTTTAAACATCTC 60


```

Y 1637 GTTCGTCATCTCCACCGTCGGTTTCTCTCAGCTATATTTA 1677
b 61 GTTCGTCATCTCCACCGTCGGTTTCTCTCAGCTATATTTA 101

RESULT 8
BES23004
LOCUS
DEFINITION
M31B6STM Arabidopsis developing seed Arabidopsis thaliana cDNA
clone M31B6 5', mRNA sequence.
ACCESSION
BES23004
VERSION
BES23004.1 GI:9780982
KEYWORDS
EST
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 378)
AUTHORS
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL
Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE
20567808
PUBMED
11115876
COMMENT
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.

FEATURES
Location/Qualifiers
1..378
/molecule="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="M31B6"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_lib="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site 1: EcoRI; Site 2: XhoII"
BASE COUNT 108 a 97 c 81 g 92 t
ORIGIN
1..378
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="M31B6"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_lib="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site 1: EcoRI; Site 2: XhoII"
Query Match 5.1%; Score 85; DB 10; Length 378;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1593 GTTAAACGATCTTTCCTCCATGTCATCGCTTTTAACTCTCGTCATCTCCACCG 1652
b 18 GTTAAACGATCTTTCCTCCATGTCATCGCTTTTAACTCTCGTCATCTCCACCG 77

RESULT 9
BES23004
LOCUS
DEFINITION
SALK_035853 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_035853, genomic survey sequence.
ACCESSION
BES23004
VERSION
BES23004.1 GI:9780982
KEYWORDS
EST
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 146)
AUTHORS
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.
TITLE
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL
Unpublished
COMMENT
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within 300 bases of the 5' end of
At2G45680.
Class: TDNA tagged.
Location/Qualifiers
1..146
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
BASE COUNT 33 a 44 c 27 g 42 t
ORIGIN
1..146
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
Query Match 4.9%; Score 82; DB 28; Length 146;
Best Local Similarity 100.0%; Pred. No. 4e-30;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1596 AAAACGATCTTTCCTCCATGTCATCGCTTTTAACTCTCGTCATCTCCACCGTCC 1655
b 1 AAAACGATCTTTCCTCCATGTCATCGCTTTTAACTCTCGTCATCTCCACCGTCC 60

Y 1656 GTTTTCTCTCAGCTATATTTA 1677
b 61 GTTTTCTCTCAGCTATATTTA 82

RESULT 10
CC459772/c
LOCUS
DEFINITION
CC459772 600 bp DNA linear GSS 03-JUN-2003
SALK_133376.34.05.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_133376.34.05.x, genomic
survey sequence.
ACCESSION
CC459772
VERSION
CC459772.1 GI:31343739
KEYWORDS
GSS
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 600)
AUTHORS
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 146)
AUTHORS
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.
TITLE
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL
Unpublished
COMMENT
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within 300 bases of the 5' end of
At2G45680.
Class: TDNA tagged.
Location/Qualifiers
1..146
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
BASE COUNT 33 a 44 c 27 g 42 t
ORIGIN
1..146
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
Query Match 4.9%; Score 82; DB 28; Length 146;
Best Local Similarity 100.0%; Pred. No. 4e-30;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1596 AAAACGATCTTTCCTCCATGTCATCGCTTTTAACTCTCGTCATCTCCACCGTCC 1655
b 1 AAAACGATCTTTCCTCCATGTCATCGCTTTTAACTCTCGTCATCTCCACCGTCC 60

Y 1656 GTTTTCTCTCAGCTATATTTA 1677
b 61 GTTTTCTCTCAGCTATATTTA 82

RESULT 10
CC459772/c
LOCUS
DEFINITION
CC459772 600 bp DNA linear GSS 03-JUN-2003
SALK_133376.34.05.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_133376.34.05.x, genomic
survey sequence.
ACCESSION
CC459772
VERSION
CC459772.1 GI:31343739
KEYWORDS
GSS
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 600)
AUTHORS
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrihab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
Zimmerman,J. and Ecker,J.R.

```

TITLE A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL Unpublished
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.

Location/Qualifiers

FEATURES

source

1. .600

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_133376_34_05_x"

/clone_lib="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/tdna_protocols.html"

160 a 130 c 102 g 166 t 42 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

40; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

2y

1057

TAATGGGTCCACCCCAAGTATGGCTTACACTTTTC 1096

|||||

475

TAATGGGTCCACCCCAAGTATGGGTACACTTTTC 436

|||||

RESULT 11

BH493162/c

LOCUS

BH493162

DEFINITION

BOGOH91TR BOGO Brassica oleracea genomic clone BOGOH91, genomic

survey sequence.

ACCSSION

BH493162

VERSION

BH493162.1

KEYWORDS

GSS

SOURCE

Brassica oleracea

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 825)

Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .825

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOGOH91"

/clone_lib="BOGO"

source

BASE COUNT
ORIGIN

Query Match

Best Local Similarity

Matches

26; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

1058

AATGGGTCCACCCCAAGTATGGC 1083

|||||

Db

91

AATGGGTCCACCCCAAGTATGGC 66

|||||

RESULT 12

AL764163

LOCUS

AL764163

DEFINITION

Arabidopsis thaliana T-DNA flanking sequence GK-122B06-012551,

genomic survey sequence.

ACCSSION

AL764163

VERSION

AL764163.1

KEYWORDS

GSS

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1

Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H.

and Weisshaar, B.

A pipeline for automated high-throughput generation of FSTs

(flanking sequence tags) from Arabidopsis thaliana T-DNA

transformed lines

Unpublished

2

Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)

for flanking sequence tag based reverse genetics

Unpublished

3 (bases 1 to 207)

Li, Y., Rosso, M., Strizhov, N. and Weisshaar, B.

Direct Submission

Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer

Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

This sequence is recovered from the left border of the T-DNA. It

indicates an insertion within the locus defined by clone F4H6. The

sequences are generated at the MPI for Plant Breeding Research in

the context of the GABI-Kat project. GABI-Kat is part of the German

Plant Genomics program designated 'GABI'. Information on line

availability can be found at:

<http://www.mpiz-koein.mpg.de/GABI-Kat/>.

Location/Qualifiers

1. .207

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-122B06-012551"

/clone_lib="Arabidopsis thaliana T-DNA insertion lines"

/note="PCR was performed on DNA from Arabidopsis thaliana

plants (T1) which were transformed with the T-DNA from

vector PAC161. The lines contain one or more T-DNA

insertions. The DNA fragment(s) resulting from the PCR

were directly sequenced to determine the genomic sequence

flanking the insertion. Sequences displaying significant

similarity to the A. thaliana nuclear genome sequence were

processed for submission. T-DNA derived sequences were

removed"

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

100.0%;

Pred. No. 16;

Score 22;

Length 207;

/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

292 a 174 c 137 g 222 t

Query Match

Best Local Similarity

Matches

26; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

1058

AATGGGTCCACCCCAAGTATGGC 1083

|||||

Db

91

AATGGGTCCACCCCAAGTATGGC 66

|||||

RESULT 12

AL764163

LOCUS

AL764163

DEFINITION

Arabidopsis thaliana T-DNA flanking sequence GK-122B06-012551,

genomic survey sequence.

ACCSSION

AL764163

VERSION

AL764163.1

KEYWORDS

GSS

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1

Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H.

and Weisshaar, B.

A pipeline for automated high-throughput generation of FSTs

(flanking sequence tags) from Arabidopsis thaliana T-DNA

transformed lines

Unpublished

2

Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)

for flanking sequence tag based reverse genetics

Unpublished

3 (bases 1 to 207)

Li, Y., Rosso, M., Strizhov, N. and Weisshaar, B.

Direct Submission

Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer

Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

This sequence is recovered from the left border of the T-DNA. It

indicates an insertion within the locus defined by clone F4H6. The

sequences are generated at the MPI for Plant Breeding Research in

the context of the GABI-Kat project. GABI-Kat is part of the German

Plant Genomics program designated 'GABI'. Information on line

availability can be found at:

<http://www.mpiz-koein.mpg.de/GABI-Kat/>.

Location/Qualifiers

1. .207

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-122B06-012551"

/clone_lib="Arabidopsis thaliana T-DNA insertion lines"

/note="PCR was performed on DNA from Arabidopsis thaliana

plants (T1) which were transformed with the T-DNA from

vector PAC161. The lines contain one or more T-DNA

insertions. The DNA fragment(s) resulting from the PCR

were directly sequenced to determine the genomic sequence

flanking the insertion. Sequences displaying significant

similarity to the A. thaliana nuclear genome sequence were

processed for submission. T-DNA derived sequences were

removed"

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

100.0%;

Pred. No. 16;

Score 22;

Length 207;

```

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

>y 1308 ATTACTAAACAATAAAGAAAA 1329
|||||
>b 141 ATTACTAAACAATAAAGAAAA 162
|||||

RESULT 13
LOCUS AQ055392 162 bp DNA linear GSS 30-JUL-1998
DEFINITION CIT-HSP-2344P6, TR CIT-HSP Homo sapiens genomic clone 2344P6,
genomic survey sequence.
ACCESSION AQ055392
VERSION AQ055392.1 GI:3351998
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162)
Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,
Simon, M. and Venter, J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished
COMMENT Other GSSs: CIT-HSP-2344P6.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
source
1..162
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2344P6"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelosAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 56 a 36 c 22 g 48 t
ORIGIN
Query Match 1.3%; Score 21; DB 28; Length 162;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

>y 1093 TTTCATTAATAAATAAGTAAA 1113
|||||
>b 103 TTTCATTAATAAATAAGTAAA 123
|||||

RESULT 14
LOCUS BE933474/c 178 bp mRNA linear EST 02-OCT-2000
DEFINITION RC4-HT0887-250800-012-d07 HT0887 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE933474
VERSION BE933474.1 GI:10459550
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

1 (bases 1 to 178)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
MEDLINE 10737800
PUBMED
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=kt2=RC4-HT0887-250
800-012-d07&tl3=2000-08-25&tl4=1)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 154.
FEATURES
source
1..178
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0887"
/note="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 40 a 54 c 45 g 39 t
ORIGIN
Query Match 1.3%; Score 21; DB 10; Length 178;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

>y 416 TTATCCCTCTGTGTGGACCC 436
|||||
>b 43 TTATCCCTCTGTGTGGACCC 23
|||||

RESULT 15
LOCUS BF703799 258 bp mRNA linear EST 22-DEC-2000
DEFINITION MI-P-E4-abq-h-11-1-UM.s1 MI-P-E4 Sus scrofa cDNA clone
MI-P-E4-abq-h-11-1-UM 3', mRNA sequence.
ACCESSION BF703799
VERSION BF703799.1 GI:11989207
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 258)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 97044477
MEDLINE 8889548
PUBMED
Contact: Tuggle CK

```

Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized embryo at gestational day 14 library cDNA library
Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
Research Center, Department of Animal Science, University of
Missouri-Columbia, 65211 Clone distribution: clones will be
available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward
POLYA=Yes.

FEATURES Location/Qualifiers
source 1..258
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /strain="crossbreed"
 /db_xref="taxon:9823"
 /clone="MI-P-E4-abq-h-11-1-UM"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="MI-P-E4"
 /notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: EcoRI; The MI-P-E4
library is derived from embryo at gestational day 14. For
a detailed description of the library from which this
clone was derived, please visit our web site at
http://pigest.genome.iastate.edu/.

TAG LIB=MI-P-E4
TAG TISSUE=embryo at gestational day 14
TAG_SEQ=ACTCAC

BASE COUNT 79 a 52 g 96 t
ORIGIN
Query Match 1.3%; Score 21; DB 10; Length 258;
Best Local Similarity 100.0%; Pred.No. 51;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 284 TGCGTTAGTGAAGTAAA 304
Db 150 TGCGTTAGTGAAGTAAA 170

Search completed: February 2, 2004, 05:01:24
Job time : 3888.05 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: February 2, 2004, 02:56:51 ; Search time 111.678 Seconds
(without alignments)
6627.983 Million cell updates/sec

itle: US-09-938-842A-3729

erfect score: 1677
equene: 1 ggtaagcgtttactatg.....ttctctcagctatattta 1677

oring table: OLIGO NUC
Gapo 60.0 , Gapext 60.0

earched: 569978 seqs, 220691566 residues

ord size : 0

otal number of hits satisfying chosen parameters: 1139956

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Listing first 45 summaries

atabase : Issued Patents NA.*

- 1: /cgm2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgm2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgm2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgm2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgm2_6/ptodata/2/ina/PTCUTS_COMB.seq.*
- 6: /cgm2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result	No.	Score	Query Match	Length	ID	Description
1	18	1.1	832	1	US-08-340-539A-6	Sequence 6, Appli
2	18	1.1	832	2	US-08-461-592B-6	Sequence 6, Appli
3	18	1.1	876	4	US-09-198-119C-48	Sequence 48, Appl
4	18	1.1	877	4	US-09-198-119C-72	Sequence 72, Appl
5	18	1.1	887	4	US-09-198-119C-68	Sequence 68, Appl
6	18	1.1	1028	3	US-09-249-180-1	Sequence 1, Appli
7	18	1.1	1132	4	US-09-198-119C-56	Sequence 56, Appl
8	18	1.1	1163	3	US-09-249-180-5	Sequence 5, Appli
9	18	1.1	2148	4	US-09-328-352-2704	Sequence 2704, Ap
10	18	1.1	580073	4	US-08-545-528D-1	Sequence 1, Appli
11	18	1.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
12	18	1.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
13	17	1.0	28	1	US-08-120-827-71	Sequence 71, Appl
14	17	1.0	28	1	US-08-478-675-71	Sequence 71, Appl
15	17	1.0	210	4	US-09-134-001C-2588	Sequence 2588, Ap
16	17	1.0	241	4	US-09-389-681-400	Sequence 400, App
17	17	1.0	241	4	US-09-620-405B-400	Sequence 400, App
18	17	1.0	241	4	US-09-433-886B-400	Sequence 400, App
19	17	1.0	241	4	US-09-604-287A-400	Sequence 400, App
20	17	1.0	267	4	US-09-313-294A-1630	Sequence 1630, Ap
21	17	1.0	326	1	US-08-525-596B-7	Sequence 7, Appli
22	17	1.0	326	3	US-09-177-860A-7	Sequence 7, Appli
23	17	1.0	326	4	US-09-378-238-7	Sequence 7, Appli
24	17	1.0	326	4	US-09-451-501-7	Sequence 7, Appli
25	17	1.0	326	4	US-09-629-938-7	Sequence 7, Appli
26	17	1.0	835	4	US-09-171-209-42	Sequence 42, Appli
27	17	1.0	938	4	US-09-843-472-2	Sequence 2, Appli

28 17 1.0 1128 4 US-09-252-149B-1 Sequence 1, Appli
29 17 1.0 1128 4 US-09-451-501-20 Sequence 20, Appl
30 17 1.0 1128 4 US-09-451-501-26 Sequence 26, Appl
31 17 1.0 1196 3 US-08-891-789B-1 Sequence 1, Appli
32 17 1.0 1240 3 US-08-891-789B-3 Sequence 3, Appli
33 17 1.0 1326 3 US-09-100-391-1 Sequence 1, Appli
34 17 1.0 1326 4 US-09-616-614-1 Sequence 1, Appli
35 17 1.0 1746 4 US-09-107-532A-1731 Sequence 1731, Ap
36 17 1.0 1748 4 US-09-620-312D-60 Sequence 60, Appl
37 17 1.0 1952 2 US-08-481-337A-7 Sequence 7, Appli
38 17 1.0 1952 3 US-09-382-256-17 Sequence 17, Appl
39 17 1.0 1952 4 US-08-436-285-17 Sequence 17, Appl
40 17 1.0 1952 4 US-09-679-187-17 Sequence 17, Appl
41 17 1.0 1952 5 PCT-US95-05467-7 Sequence 7, Appli
42 17 1.0 1990 4 US-08-961-527-232 Sequence 232, App
43 17 1.0 2076 3 US-08-123-934A-3 Sequence 3, Appli
44 17 1.0 2076 5 PCT-US94-10080-3 Sequence 3, Appli
45 17 1.0 2076 5 PCT-US94-10080-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-340-539A-6
Sequence 6, Application US/08340539A
Patent No. 5808025
GENERAL INFORMATION:
APPLICANT: Tedder, Thomas F.
APPLICANT: Kansas, Geoffrey S.
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSER: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,539A
FILING DATE: 16-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,459
FILING DATE: 25-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gumison, Jane
REGISTRATION NUMBER: 38,479
REFERENCE/DOCKET NUMBER: CG-104 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-340-539A-6

Query Match: 1.1%; Score 18; DB 1; Length 832;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 TCTATCTCTAAAGAAATG 206

db 86 TCTATGCTAAAGAAATG 103
|||||

RESULT 2
US-08-461-592B-6
; Sequence 6, Application US/08461592B
; Patent No. 5834425
; GENERAL INFORMATION:
; APPLICANT: Tedder, Thomas F.
; APPLICANT: Kansas, Geoffrey S.
; TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
; TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,592B
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/340,539
; FILING DATE: 16-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/008,459
; FILING DATE: 25-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: James F. Haley, Jr.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: CG-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-461-592B-6

Query Match 1.1%; Score 18; DB 2; Length 832;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 TCTATGCTAAAGAAATG 206
DB 86 TCTATGCTAAAGAAATG 103
|||||

RESULT 3
US-09-198-119C-48/c
; Sequence 48, Application US/09198119C
; Patent No. 6417428
; GENERAL INFORMATION:
; APPLICANT: Thomasow, Michael
; APPLICANT: Stockinger, Eric
; APPLICANT: Jaglo-Ottosen, Kirsten
; APPLICANT: Gilmour, Sarah

; APPLICANT: Zarka, Daniel
; APPLICANT: Jiang, Cai-Zhong
; TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance
; FILE REFERENCE: 19117.713 Seq List
; CURRENT APPLICATION NUMBER: US/09/198,119C
; CURRENT FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: US 08/706,270
; PRIOR FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: US 09/018,233
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/017,816
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/018,235
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/017,575
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/018,227
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/018,234
; PRIOR FILING DATE: 1998-02-03
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 48
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; OTHER INFORMATION: bncBF2 gene
US-09-198-119C-48

Query Match 1.1%; Score 18; DB 4; Length 876;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1364 CCCATTGTACAGATGCT 1381
DB 815 CCCATTGTACAGATGCT 798
|||||

RESULT 4
US-09-198-119C-72/c
; Sequence 72, Application US/09198119C
; Patent No. 6417428
; GENERAL INFORMATION:
; APPLICANT: Thomasow, Michael
; APPLICANT: Stockinger, Eric
; APPLICANT: Jaglo-Ottosen, Kirsten
; APPLICANT: Gilmour, Sarah
; APPLICANT: Zarka, Daniel
; APPLICANT: Jiang, Cai-Zhong
; TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance
; FILE REFERENCE: 19117.713 Seq List
; CURRENT APPLICATION NUMBER: US/09/198,119C
; CURRENT FILING DATE: 1998-11-23
; PRIOR APPLICATION NUMBER: US 08/706,270
; PRIOR FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: US 09/018,233
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/017,816
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/018,235
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/017,575
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/018,227
; PRIOR FILING DATE: 1998-02-03
; PRIOR APPLICATION NUMBER: US 09/018,234
; PRIOR FILING DATE: 1998-02-03
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 72
; LENGTH: 877
; TYPE: DNA

ORGANISM: Brassica oleracea
FEATURE:
OTHER INFORMATION: boCBF5 gene
US-09-198-119C-72

Query Match 1.1%; Score 18; DB 4; Length 877;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1364 CCCATTGTACAGATGGT 1381
|||
b 815 CCCATTGTACAGATGGT 798
|||

RESULT 5

US-09-198-119C-68/c
Sequence 68, Application US/09198119C
Patent No. 6417428
GENERAL INFORMATION:

APPLICANT: Thomashow, Michael
APPLICANT: Stockinger, Eric
APPLICANT: Jaglo-Ottosen, Kirsten
APPLICANT: Gilmore, Sarah
APPLICANT: Zarka, Daniel
APPLICANT: Jiang, Cai-Zhong
TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance
FILE REFERENCE: 19117.713 Seq List
CURRENT APPLICATION NUMBER: US/09198,119C
CURRENT FILING DATE: 1998-11-23
PRIOR APPLICATION NUMBER: US 08/706,270
PRIOR FILING DATE: 1998-09-04
PRIOR APPLICATION NUMBER: US 09/018,233
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,816
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,235
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,575
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,227
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,234
PRIOR FILING DATE: 1998-02-03
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 68
LENGTH: 887
TYPE: DNA
ORGANISM: Brassica oleracea
FEATURE:
OTHER INFORMATION: boCBF3 gene
US-09-198-119C-68

Query Match 1.1%; Score 18; DB 4; Length 887;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1364 CCCATTGTACAGATGGT 1381
|||
b 822 CCCATTGTACAGATGGT 805
|||

RESULT 6

US-09-249-180-1/c
Sequence 1, Application US/09249180
Patent No. 6268548
GENERAL INFORMATION:
APPLICANT: Elthon, Thomas E
APPLICANT: Lund, Adrian A
APPLICANT: Bhattacharjee, Dinakar
APPLICANT: Rhoads, David M.
TITLE OF INVENTION: Isolation and Characterization of Heat Shock Protein
FILE REFERENCE: UNV92819

CURRENT APPLICATION NUMBER: US/09/249,180
CURRENT FILING DATE: 1999-02-12
EARLIER APPLICATION NUMBER: 60/076/014
EARLIER FILING DATE: 1998-02-26
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Created in PatentIn Ver. 2.0, Edited in WordPerfect 6.1
SEQ ID NO 1
LENGTH: 1028
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (79)..(735)
FEATURE:
NAME/KEY: Gene
LOCATION: (1)..(1028)
FEATURE:
NAME/KEY: misc feature
LOCATION: (424)..(693)
FEATURE:
NAME/KEY: exon
LOCATION: (1)..(328)
FEATURE:
NAME/KEY: exon
LOCATION: (329)..(1028)
FEATURE:
NAME/KEY: Poly A site
LOCATION: (1028)..(1028)
FEATURE:
NAME/KEY: mRNA
LOCATION: (1)..(1028)
FEATURE:
NAME/KEY: source
LOCATION: (1)..(1028)
OTHER INFORMATION: Zea Mays L., Line B73
FEATURE:
NAME/KEY: transit peptide
LOCATION: (79)..(213)
FEATURE:
NAME/KEY: 3'UTR
LOCATION: (736)..(1028)
FEATURE:
NAME/KEY: 5'UTR
LOCATION: (1)..(78)
PUBLICATION INFORMATION:
AUTHORS: Lund, Adrian A.
AUTHORS: Blum, Paul H.
AUTHORS: Bhattacharjee, Dinakar
AUTHORS: Elthon, Thomas E.
TITLE: Heat-Stress Response of Maize Mitochondria
JOURNAL: Plant Physiol.
VOLUME: 116
PAGES: 1097-1110
DATE: 1998-03-00
US-09-249-180-1

Query Match 1.1%; Score 18; DB 3; Length 1028;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1391 ATACTGATAGAGATAGA 1408
|||
b 880 ATACTGATAGAGATAGA 863
|||

RESULT 7

US-09-198-119C-56/c
Sequence 56, Application US/09198119C
Patent No. 6417428
GENERAL INFORMATION:
APPLICANT: Thomashow, Michael
APPLICANT: Stockinger, Eric

APPLICANT: Jaglo-Ottosen, Kirsten
APPLICANT: Gilmore, Sarah
APPLICANT: Zarka, Daniel
APPLICANT: Jiang, Cai-Zhong
TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance
FILE REFERENCE: 19117.713 Seq List
CURRENT APPLICATION NUMBER: US/09/198.119C
CURRENT FILING DATE: 1998-11-23
PRIOR APPLICATION NUMBER: US 08/706,270
PRIOR FILING DATE: 1998-09-04
PRIOR APPLICATION NUMBER: US 09/018,233
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,816
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,235
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,575
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,227
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,234
PRIOR FILING DATE: 1998-02-03
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 56
LENGTH: 1132
TYPE: DNA
ORGANISM: Brassica napus
FEATURE:
OTHER INFORMATION: bnCBF6 gene
US-09-198-119C-56

Query Match 1.1%; Score 18; DB 4; Length 1132;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1364 CCCATTGTACAGATGGT 1381
DB 851 CCCATTGTACAGATGGT 834

RESULT 8
US-09-249-180-5/c
Sequence 5, Application US/09249180
Patent No. 6268548
GENERAL INFORMATION:
APPLICANT: Elthon, Thomas E
APPLICANT: Lund, Adrian A
APPLICANT: Bhatramakki, Dinakar
APPLICANT: Rhoads, David M.
TITLE OF INVENTION: Isolation and Characterization of Heat Shock Protein
FILE REFERENCE: UNV52819
CURRENT APPLICATION NUMBER: US/09/249,180
CURRENT FILING DATE: 1999-02-12
EARLIER APPLICATION NUMBER: 60/076/014
EARLIER FILING DATE: 1998-02-26
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Created in Patent In Ver. 2.0, Edited in Wordperfect 6.1
SEQ ID NO 5
LENGTH: 1163
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (79)..(337)
FEATURE:
NAME/KEY: CDS
LOCATION: (447)..(870)
FEATURE:
NAME/KEY: exon
LOCATION: (1)..(337)
FEATURE:
NAME/KEY: exon

LOCATION: (447)..(1163)
FEATURE:
NAME/KEY: gene
LOCATION: (1)..(1163)
FEATURE:
NAME/KEY: intron
LOCATION: (329)..(463)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (559)..(828)
OTHER INFORMATION: Heat Shock Domain
FEATURE:
NAME/KEY: mRNA
LOCATION: (1)..(337)
FEATURE:
NAME/KEY: mRNA
LOCATION: (447)..(1163)
FEATURE:
NAME/KEY: Poly A site
LOCATION: (1163)..(1163)
FEATURE:
NAME/KEY: source
LOCATION: (1)..(1163)
OTHER INFORMATION: Zea Mays L., Line B73
FEATURE:
NAME/KEY: transit peptide
LOCATION: (79)..(213)
FEATURE:
NAME/KEY: 5'UTR
LOCATION: (1)..(78)
FEATURE:
NAME/KEY: 3'UTR
LOCATION: (871)..(1163)
US-09-249-180-5

Query Match 1.1%; Score 18; DB 3; Length 1163;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1391 ATACTGATAGAGATAGA 1408
DB 1015 ATACTGATAGAGATAGA 998

RESULT 9
US-09-328-352-2704/c
Sequence 2704, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 2704
LENGTH: 2148
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-2704

Query Match 1.1%; Score 18; DB 4; Length 2148;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 ATGAATGGTATGCTGATG 380
DB 1263 ATGAATGGTATGCTGATG 1246

RESULT 10
US-08-545-528D-1

Sequence 1, Application US/08545528D
Patent No. 6537773
GENERAL INFORMATION:

APPLICANT: FRASER et al.
TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment

PATENT NO. 6537773

TITLE OF INVENTION: Thereof, and Uses Thereof

FILE REFERENCE: PB193F1

CURRENT APPLICATION NUMBER: US/08/545,528D

CURRENT FILING DATE: 1995-10-19

PRIOR APPLICATION NUMBER: US 08/488,018

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: US 08/473,545

PRIOR FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 1

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 580073

TYPE: DNA

ORGANISM: Mycoplasma genitalium

JS-08-545-528D-1

Query Match 1.1%; Score 18; DB 4; Length 580073;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1332 ATTTGTGAAAGAGAAA 1349

|||||

b 355690 ATTTGTGAAAGAGAAA 355707

RESULT 11

JS-09-103-840A-2/c

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match

Best Local Similarity 1.1%; Score 18; DB 3; Length 4403765;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1636 CGTCGTCATCTCCACCGT 1653

|||||

DB 4146706 CGTCGTCATCTCCACCGT 4146689

RESULT 12

US-09-103-840A-1/c

Sequence 1, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 4411529

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match

Best Local Similarity 1.1%; Score 18; DB 3; Length 4411529;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1636 CGTCGTCATCTCCACCGT 1653

|||||

DB 4154513 CGTCGTCATCTCCACCGT 4154496

RESULT 13

US-08-120-827-71/c

Sequence 71, Application US/08120827

Patent No. 5525495

GENERAL INFORMATION:

APPLICANT: KEENE, JACK D.

APPLICANT: KING, PETER H.

APPLICANT: LEVINE, TODD

TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS

TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION

NUMBER OF SEQUENCES: 101

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/120,827

FILING DATE: 15-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5525495man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 714-158-0 CIP

TELEPHONE: (703)413-3000

TELEFAX: (703)413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 71:

SEQUENCE CHARACTERISTICS:

LENGTH: 28 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: RNA (genomic)

US-08-120-827-71

Query Match

Best Local Similarity 1.0%; Score 17; DB 1; Length 28;

Matches 100.0%; Pred. No. 95;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1315 AACAAATAAAGAAAAAC 1331
Db 28 AACAAATAAAGAAAAAC 12

RESULT 14

US-08-478-675-71/c
; Sequence 71, Application US/08478675
; Patent No. 5773246
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,675
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/120,827
; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5773246man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: RNA (genomic)
US-08-478-675-71

Query Match 1.0%; Score 17; DB 1; Length 28;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1315 AACAAATAAAGAAAAAC 1331
Db 28 AACAAATAAAGAAAAAC 12

RESULT 15

US-09-134-001C-2588/c
; Sequence 2588, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2588
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2588

Query Match 1.0%; Score 17; DB 4; Length 210;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 AAACAAGCATCATTAAA 237
Db 95 AAACAAGCATCATTAAA 79

Search completed: February 2, 2004, 07:59:13
Job time : 134.678 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 03:00:32 ; Search time 1254.7 Seconds
(without alignments)
4871.226 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 ggtaacggttttacttatg.....ttctctcagctatatattta 1677

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2434939 seqs, 1822278265 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1677	100.0	1677	10	US-09-938-842A-3729
2	1677	100.0	1677	12	Sequence 3729, Ap
3	21	1.3	330	13	US-09-938-842A-3729
4	21	1.3	394	11	Sequence 54, Appl
5	21	1.3	407	11	US-09-803-719-707
6	21	1.3	2000	13	Sequence 1633, Ap
7	21	1.3	108359	15	US-09-930-213-306
8	20	1.2	627	13	US-10-191-807-3
9	20	1.2	627	13	US-10-027-632-54726
10	20	1.2	627	14	US-10-027-632-294129
11	20	1.2	627	14	Sequence 54726, A
12	20	1.2	768	9	US-10-027-632-54726
13	19	1.1	199	13	US-09-815-242-9987
14	19	1.1	281	12	Sequence 9987, Ap
15	19	1.1	377	10	US-10-029-386-14103
					Sequence 14103, A
					Sequence 18166, A
					Sequence 5678, Ap

US-09-938-842A-3729
Sequence 3729, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Krops, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US/60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US/60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US/60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO: 3729
LENGTH: 1677
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-3729

Query Match 100.0%; Score 1677; DB 10; Length 1677;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTTAAGCGTTTACTTATGTTATGCAACGGAGATATGCAATGTTGGATGC 60
DB 1 GGTTAAGCGTTTACTTATGTTATGCAACGGAGATATGCAATGTTGGATGC 60
QY 61 TTTTTCAGATCATCAAGGCTTCCTACAGATTTCTAGGATGTTTCAGGCTTTTGTTA 120
DB 61 TTTTTCAGATCATCAAGGCTTCCTACAGATTTCTAGGATGTTTCAGGCTTTTGTTA 120

ALIGNMENTS

RESULT 1

US-09-938-842A-3729
Sequence 3729, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Krops, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US/60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US/60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US/60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO: 3729
LENGTH: 1677
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-3729

121 GAAATGTTGTTTATTCGCAACAGGTAGAGAAATACCAATAGACAGATGATCTGAAGAGA 180
121 GAAATGTTGTTTATTCGCAACAGGTAGAGAAATACCAATAGACAGATGATCTGAAGAGA 180
181 TAAGCTTCTATGCTAAAGAAATGGACCGATAGCAATTAACCAAGCATCATTAAGAT 240
181 TAAGCTTCTATGCTAAAGAAATGGACCGATAGCAATTAACCAAGCATCATTAAGAT 240
241 TAAATGGTTTGAAGAAATACCTATCTATTTATGTAATTTGCTGCTAGTGAAGAT 300
241 TAAATGGTTTGAAGAAATACCTATCTATTTATGTAATTTGCTGCTAGTGAAGAT 300
301 AAAACATCGGAATCCAAACCTCCTAAATTTACCAATCAGCCCAATTAATGATGCTGCG 360
301 AAAACATCGGAATCCAAACCTCCTAAATTTACCAATCAGCCCAATTAATGATGCTGCG 360
361 TAAATGAATGCTATGCTAGTGAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
361 TAAATGAATGCTATGCTAGTGAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
421 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
421 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
481 AAGCCTAAAGGCTTACTACCGGTTTGAACCGTTTATATTTGGTGTAAATCTTAATCCC 540
481 AAGCCTAAAGGCTTACTACCGGTTTGAACCGTTTATATTTGGTGTAAATCTTAATCCC 540
541 GGATCCGTTTGTGTTTAACTCAAGGCCACGTTATCGCCCAATTTTGAATTTTGAGTG 600
541 GGATCCGTTTGTGTTTAACTCAAGGCCACGTTATCGCCCAATTTTGAATTTTGAGTG 600
601 GGTAGGGAATGTTGGGTCGAATAGTGGGCTAGCCCTCAACAAATGTTGGAACTGAAG 660
601 GGTAGGGAATGTTGGGTCGAATAGTGGGCTAGCCCTCAACAAATGTTGGAACTGAAG 660
661 AGAGTAGGCTCAGCTCAGGCCCAATTCACATTTTCTGTTTGTAGCCCTTTCTTTCTCG 720
661 AGAGTAGGCTCAGCTCAGGCCCAATTCACATTTTCTGTTTGTAGCCCTTTCTTTCTCG 720
721 TGCTTACGCTCCCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
721 TGCTTACGCTCCCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
781 CCGGAACAGTACCAAGAAATCAAAATAGTTTGAATCGTTTCAATCTAGTACCGTGC 840
781 CCGGAACAGTACCAAGAAATCAAAATAGTTTGAATCGTTTCAATCTAGTACCGTGC 840
841 AACTTACAAATCTATTCGATTTCTGATCTGATTTCTAGTTCTGATTTCTGATTTCTAT 900
841 AACTTACAAATCTATTCGATTTCTGATCTGATTTCTAGTTCTGATTTCTGATTTCTAT 900
901 CCGGATTGTACAGTACCAAGTACATAGTATGCTATATGATGATGATGATGATGATGATG 960
901 CCGGATTGTACAGTACCAAGTACATAGTATGCTATATGATGATGATGATGATGATGATG 960
961 AAAGCAGCAGCATAGGAGGATTTTGGAAATCCTGGAAGAGGATTTTCCATAGACACTA 1020
961 AAAGCAGCAGCATAGGAGGATTTTGGAAATCCTGGAAGAGGATTTTCCATAGACACTA 1020
1021 ATTAGCTTTTGGTGGCGCAGCTGTGACCTTACATTAATGAGGCTGCAACCCCAAGTATG 1080
1021 ATTAGCTTTTGGTGGCGCAGCTGTGACCTTACATTAATGAGGCTGCAACCCCAAGTATG 1080
1081 GCTTACAGCTTTTCCATAAATTAAGTAAATCTTTTGGCTTACCAATTAAGTAAAT 1140
1081 GCTTACAGCTTTTCCATAAATTAAGTAAATCTTTTGGCTTACCAATTAAGTAAAT 1140
1141 ATTGAAATCTTTTCCAAACATAGAAAGTAAATTTGATCAGCGATGGAATTTTGTGAC 1200
1141 ATTGAAATCTTTTCCAAACATAGAAAGTAAATTTGATCAGCGATGGAATTTTGTGAC 1200

1201 AAAGCTAGGTATTTCAATTTGGAGGTGCTACTAGTAAGTACTAGTAAGTACTAGTAAGT 1260
1201 AAAGCTAGGTATTTCAATTTGGAGGTGCTACTAGTAAGTACTAGTAAGTACTAGTAAGT 1260
1261 TTCTGATTTTGGATTTTGAAGCTTTTCTAGGTTAAAGAAAGGATTAAGTACTAAGCAAT 1320
1261 TTCTGATTTTGGATTTTGAAGCTTTTCTAGGTTAAAGAAAGGATTAAGTACTAAGCAAT 1320
1321 AAAGGAAAAACATTTTGTGAAAGAGAAATAAGTTTACTGAGCCCAATTTGTACAGATGG 1380
1321 AAAGGAAAAACATTTTGTGAAAGAGAAATAAGTTTACTGAGCCCAATTTGTACAGATGG 1380
1381 TCCCATTAATAACTGATAGAGATAGAGCAATGGAAGTGAATTTGTTCACGTGTACAA 1440
1381 TCCCATTAATAACTGATAGAGATAGAGCAATGGAAGTGAATTTGTTCACGTGTACAA 1440
1441 TCGGAATGGTTCTTTAAAGCTCATCGAACACATCAGGACCGTTGATTTTCCCGCATCAA 1500
1441 TCGGAATGGTTCTTTAAAGCTCATCGAACACATCAGGACCGTTGATTTTCCCGCATCAA 1500
1501 AAAGCGTTGAATTAATTTCTACTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1501 AAAGCGTTGAATTAATTTCTACTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1561 ACATTTAGTAATCTCTGAGCTGTAAGCGGTTAAAGCAATTTTCCCAATTTGATCC 1620
1561 ACATTTAGTAATCTCTGAGCTGTAAGCGGTTAAAGCAATTTTCCCAATTTGATCC 1620
1621 GCTTTTAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677
1621 GCTTTTAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677

RESULT 2

US-09-938-842A-3729
; Sequence 3729, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,966
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3729
; LENGTH: 1677
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3729

Query Match 100.0%; Score 1677; DB 12; Length 1677;
Best Local Similarity 100.0%; Freq. No. 0;
Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATGCCCATTTGTTGGAATGC 60
DB 1 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATGCCCATTTGTTGGAATGC 60
QY 61 TTTTCAGATCATCAAGCTCTACAGATTTCTTAGGAATGTTTCAGGCTTTGTTA 120
DB 61 TTTTCAGATCATCAAGCTCTACAGATTTCTTAGGAATGTTTCAGGCTTTGTTA 120
QY 121 GAAATGTGTTTATTGCAACAGGTAGAGAACATAACCATAGACAGATCTATCTGAAGAGA 180

Db 121 GAAATGTTTATTTGCAACAGGTAGAGAAATACCAATAGACAGATGTATCTGAAGAGA 180
Qy 181 TAAGCTTCTCTATGCTTAAGAAATGGAACCGATACGAATAAAAACAAGCATCATTAAGAT 240
Db 181 TAAGCTTCTCTATGCTTAAGAAATGGAACCGATACGAATAAAAACAAGCATCATTAAGAT 240
Qy 241 TAAATGGTTTGTAGAAATACCTACACTTATTTATGTGAATTTGTGGTTAGTGAAGAAT 300
Db 241 TAAATGGTTTGTAGAAATACCTACACTTATTTATGTGAATTTGTGGTTAGTGAAGAAT 300
Qy 301 AAAAATCATCGGAATCCAAAATCTCAAAATTTACCAATCAGCCCAATTAATGTGCTGGCG 360
Db 301 AAAAATCATCGGAATCCAAAATCTCAAAATTTACCAATCAGCCCAATTAATGTGCTGGCG 360
Qy 361 TAAATGAATGGTATGCTGATGTAGTGAAGTGGTGGTGGCGGAAATTAACAATATTC 420
Db 361 TAAATGAATGGTATGCTGATGTAGTGAAGTGGTGGTGGCGGAAATTAACAATATTC 420
Qy 421 CCTCTGTGGTGGACCGCAATCTGTAATCGGAAGAGTGGACCCACCTGGTTTAACTTTT 480
Db 421 CCTCTGTGGTGGACCGCAATCTGTAATCGGAAGAGTGGACCCACCTGGTTTAACTTTT 480
Qy 481 AAGCCTAAAAGGTTACTACCGGTTTGACCGGTTTATTAATTTGGTGTAAATCTTAATCCC 540
Db 481 AAGCCTAAAAGGTTACTACCGGTTTGACCGGTTTATTAATTTGGTGTAAATCTTAATCCC 540
Qy 541 GGATCCGTTTGGTTTAAATCTCAAGCCACGTTATCGCCAAATTTTGAATTTTGTAGTG 600
Db 541 GGATCCGTTTGGTTTAAATCTCAAGCCACGTTATCGCCAAATTTTGAATTTTGTAGTG 600
Qy 601 GGTAGGGAATCGTGGGTGCAATAGTTGGGCTAGCCCTCAACCAATGTGGAACTGAAG 660
Db 601 GGTAGGGAATCGTGGGTGCAATAGTTGGGCTAGCCCTCAACCAATGTGGAACTGAAG 660
Qy 661 AGATGAGGTTCCAGCTCAGGCCACATTCATTTTGGTGTAGTGGTGTAGTGGTGTAGT 720
Db 661 AGATGAGGTTCCAGCTCAGGCCACATTCATTTTGGTGTAGTGGTGTAGTGGTGTAGT 720
Qy 721 TGCTTACGGTCCCTTCTGCTGGTGGTATGTAAGTGAATCGGTTACATCTAGTTACCGTCG 780
Db 721 TGCTTACGGTCCCTTCTGCTGGTGGTATGTAAGTGAATCGGTTACATCTAGTTACCGTCG 780
Qy 781 CCCGAAACAGTACCAAGATCAAAATAGTTTGAATCGGTTACATCTAGTTACCGTCG 840
Db 781 CCCGAAACAGTACCAAGATCAAAATAGTTTGAATCGGTTACATCTAGTTACCGTCG 840
Qy 841 AACTTCAATCATTTGATTTGATCTGATTTCTGATTTCTAGTTTGGTGTAGTTTGAATAT 900
Db 841 AACTTCAATCATTTGATTTGATCTGATTTCTGATTTCTAGTTTGGTGTAGTTTGAATAT 900
Qy 901 CCGGATTTCAAGTACACAAGTACATAAGTATCGGTATATGATGTGACCGGTTTAAATC 960
Db 901 CCGGATTTCAAGTACACAAGTACATAAGTATCGGTATATGATGTGACCGGTTTAAATC 960
Qy 961 AAAGACGACGATAGGAGGATTTGGAATCTCGGAAGAGGATTAATTCATAGACACTA 1020
Db 961 AAAGACGACGATAGGAGGATTTGGAATCTCGGAAGAGGATTAATTCATAGACACTA 1020
Qy 1021 ATTAGCTTTTGGTGGCCACCGCTTGACCTACATTAATGGGTCCACCCCAAGTATG 1080
Db 1021 ATTAGCTTTTGGTGGCCACCGCTTGACCTACATTAATGGGTCCACCCCAAGTATG 1080
Qy 1081 GCGTTACAGCTTTTCCATAAAATTAAGTAAATCTTTTGGCTAACCAATAAATAAT 1140
Db 1081 GCGTTACAGCTTTTCCATAAAATTAAGTAAATCTTTTGGCTAACCAATAAATAAT 1140
Qy 1141 ATTGAAATCTTCCACCAATAGAAAGTAAATTTGATCAGCATCGAAATTTTGTATC 1200
Db 1141 ATTGAAATCTTCCACCAATAGAAAGTAAATTTGATCAGCATCGAAATTTTGTATC 1200
Qy 1201 AAAGCTAGTATTTTCAATTTGGAGTGTACTAGTAACTAGTAACTAAACAGATGAGT 1260

Db 1201 AAAGCTAGTATTTTCAATTTGGAGTGTACTAGTAACTAGTAACTAAACAGATGAGT 1260
Qy 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTACTAAACAAT 1320
Db 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTACTAAACAAT 1320
Qy 1321 AAAAGAAAACATTTTGTGAAAAGAGAAATAAAGTTTACTGGACCCCATTTGACAGATGG 1380
Db 1321 AAAAGAAAACATTTTGTGAAAAGAGAAATAAAGTTTACTGGACCCCATTTGACAGATGG 1380
Qy 1381 TCCCATATAATTAATCTAGATAGAGATAGAGCAATGGAAGTGAATTTGTTCACTGGTACAA 1440
Db 1381 TCCCATATAATTAATCTAGATAGAGATAGAGCAATGGAAGTGAATTTGTTCACTGGTACAA 1440
Qy 1441 TCGGAATGGTTCTTTTAAAGCTCATCGAACAATCAGGACCGTTGATTTTCCCGCATCAA 1500
Db 1441 TCGGAATGGTTCTTTTAAAGCTCATCGAACAATCAGGACCGTTGATTTTCCCGCATCAA 1500
Qy 1501 AAAGCGTTGAATTAATCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 1501 AAAGCGTTGAATTAATCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Qy 1561 ACATTTAGTAAATCTCCTTTGAGCGTGAACGCGTTAAACGATTTTCCCATTTGATATCC 1620
Db 1561 ACATTTAGTAAATCTCCTTTGAGCGTGAACGCGTTAAACGATTTTCCCATTTGATATCC 1620
Qy 1621 GCTTTTAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677
Db 1621 GCTTTTAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677

RESULT 3
US-09-930-213-54
; Sequence 54, Application US/09930213
; Publication No. US20030170625A1
; GENERAL INFORMATION:
; APPLICANT: ROSENTHAL, ANDRE
; APPLICANT: HINZMANN, BERNHARD
; APPLICANT: SCHAFER, REINHARD
; APPLICANT: ZUBER, JOHANNES
; APPLICANT: TCHER-NITSE, OLEG
; APPLICANT: GRIFF, MARTIN
; APPLICANT: HELLMER, MARTIN
; APPLICANT: SCHMITZ, ANNE-CHANTAL
; APPLICANT: SERS, CHRISTINE
; TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
; FILE REFERENCE: ALBRE-14
; CURRENT APPLICATION NUMBER: US/09/930,213
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: DE 10004102.7
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 885
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 54
; LENGTH: 330
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (4)
; OTHER INFORMATION: a, t, c, g, other or unknown
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (9)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-09-930-213-54

Query Match 1.3%; Score 21; DB 13; Length 330;
Best Local Similarity 100.0%; Pred.No. 7.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TTATCCCTCTGTTGGTGGACCC 436

Db 148 TTATCCCTCTGTGGTGACCC 168

RESULT 4

US-09-803-719-707

Sequence 707, Application US/09803719

Publication No. US20030044783A1

GENERAL INFORMATION:

APPLICANT: Williams, Lewis T.

APPLICANT: Escobedo, Jaime

APPLICANT: Innis, Michael A.

APPLICANT: Garcia, Pablo Dominguez

APPLICANT: Sudduth-Klinger, Julie

APPLICANT: Reinhard, Christoph

APPLICANT: Giese, Klaus

APPLICANT: Randazzo, Filippo

APPLICANT: Kennedy, Giulia C.

APPLICANT: Pot, David

APPLICANT: Kassam, Altaf

APPLICANT: Lamson, George

APPLICANT: Drmanac, Radoje

APPLICANT: Crkvenjakov, Radomir

APPLICANT: Dickson, Mark

APPLICANT: Drmanac, Snezana

APPLICANT: Labat, Ivan

APPLICANT: Leshkowitz, Dena

APPLICANT: Kita, David

APPLICANT: Garcia, Veronica

APPLICANT: Jones, Lee William

APPLICANT: Stache-Crain, Birgit

TITLE OF INVENTION: Human Genes and Gene Products

FILE REFERENCE: 1624.002

CURRENT APPLICATION NUMBER: US/09/803,719

CURRENT FILING DATE: 2001-03-09

PRIOR APPLICATION NUMBER: 60/188,609

PRIOR FILING DATE: 2000-03-09

NUMBER OF SEQ ID NOS: 2396

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 707

LENGTH: 394

TYPE: DNA

ORGANISM: Homo sapiens

US-09-803-719-707

Query Match 1.3%; Score 21; DB 11; Length 394;

Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TTATCCCTCTGTGGTGACCC 436

Db 49 TTATCCCTCTGTGGTGACCC 69

RESULT 5

US-09-803-719-1633

Sequence 1633, Application US/09803719

Publication No. US20030044783A1

GENERAL INFORMATION:

APPLICANT: Williams, Lewis T.

APPLICANT: Escobedo, Jaime

APPLICANT: Innis, Michael A.

APPLICANT: Garcia, Pablo Dominguez

APPLICANT: Sudduth-Klinger, Julie

APPLICANT: Reinhard, Christoph

APPLICANT: Giese, Klaus

APPLICANT: Randazzo, Filippo

APPLICANT: Kennedy, Giulia C.

APPLICANT: Pot, David

APPLICANT: Kassam, Altaf

APPLICANT: Lamson, George

APPLICANT: Drmanac, Radoje

APPLICANT: Crkvenjakov, Radomir

APPLICANT: Dickson, Mark

TITLE OF INVENTION: Human Genes and Gene Products

FILE REFERENCE: 1624.002

CURRENT APPLICATION NUMBER: US/09/803,719

CURRENT FILING DATE: 2001-03-09

PRIOR APPLICATION NUMBER: 60/188,609

PRIOR FILING DATE: 2000-03-09

NUMBER OF SEQ ID NOS: 2396

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 707

LENGTH: 394

TYPE: DNA

ORGANISM: Homo sapiens

US-09-803-719-707

Query Match 1.3%; Score 21; DB 11; Length 394;

Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TTATCCCTCTGTGGTGACCC 436

Db 49 TTATCCCTCTGTGGTGACCC 69

APPLICANT: Drmanac, Snezana

APPLICANT: Labat, Ivan

APPLICANT: Leshkowitz, Dena

APPLICANT: Kita, David

APPLICANT: Garcia, Veronica

APPLICANT: Jones, Lee William

APPLICANT: Stache-Crain, Birgit

TITLE OF INVENTION: Human Genes and Gene Products

FILE REFERENCE: 1624.002

CURRENT APPLICATION NUMBER: US/09/803,719

CURRENT FILING DATE: 2001-03-09

PRIOR APPLICATION NUMBER: 60/188,609

PRIOR FILING DATE: 2000-03-09

NUMBER OF SEQ ID NOS: 2396

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1633

LENGTH: 407

TYPE: DNA

ORGANISM: Homo sapiens

US-09-803-719-1633

Query Match 1.3%; Score 21; DB 11; Length 407;

Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TTATCCCTCTGTGGTGACCC 436

Db 195 TTATCCCTCTGTGGTGACCC 215

RESULT 6

US-09-930-213-306

Sequence 306, Application US/09930213

Publication No. US20030170625A1

GENERAL INFORMATION:

APPLICANT: ROSENTHAL, ANDRE

APPLICANT: HINZMANN, BERND

APPLICANT: SCHAFER, REINHARD

APPLICANT: ZUBER, JOHANNES

APPLICANT: TCHE-NITSE, OLEG

APPLICANT: GRIPS, MARTIN

APPLICANT: HELLMER, MARTIN

APPLICANT: SCHMITZ, ANNE-CHANTAL

APPLICANT: SERS, CHRISTINE

TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS

FILE REFERENCE: ALBRE-14

CURRENT APPLICATION NUMBER: US/09/930,213

CURRENT FILING DATE: 2001-01-31

PRIOR APPLICATION NUMBER: DE 10004102.7

PRIOR FILING DATE: 2000-01-31

NUMBER OF SEQ ID NOS: 885

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 306

LENGTH: 2000

TYPE: DNA

ORGANISM: Homo sapiens

US-09-930-213-306

Query Match 1.3%; Score 21; DB 13; Length 2000;

Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TTATCCCTCTGTGGTGACCC 436

Db 885 TTATCCCTCTGTGGTGACCC 905

RESULT 7

US-10-191-807-3/c

Sequence 3, Application US/10191807

Publication No. US20030068691A1

GENERAL INFORMATION:

APPLICANT: HU, Song et al.

;; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;; TITLE OF INVENTION: US\$ THEREOF
;; FILE REFERENCE: CL001273-PROV
;; CURRENT APPLICATION NUMBER: US/10/191.807
;; CURRENT FILING DATE: 2002-07-10
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 108359
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)-(108359)
;; OTHER INFORMATION: n = A,T,C or G
JS-10-191-807-3

Query Match 1.3%; Score 21; DB 15; Length 108359;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 240 TTAATGGTTTGAAGAATA 260
|||||
b 75869 TTAATGGTTTGAAGAATA 75849

RESULT 8
JS-10-027-632-54726/c
; Sequence 54726, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54726
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(627)
; OTHER INFORMATION: n = A,T,C or G
JS-10-027-632-54726

Query Match 1.2%; Score 20; DB 13; Length 627;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1134 AAAAATTATTGAAATCTTT 1153
|||||
b 508 AAAAATTATTGAAATCTTT 489

RESULT 9
US-10-027-632-294129/c
; Sequence 294129, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 294129
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(627)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-294129

Query Match 1.2%; Score 20; DB 13; Length 627;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 AAAAATTATTGAAATCTTT 1153
|||||
b 508 AAAAATTATTGAAATCTTT 489

RESULT 10
US-10-027-632-54726/c
; Sequence 54726, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54726

LENGTH: 627
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(627)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-54726

Query Match
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 AAAAATTATTGAAATCTTT 1153
DB 508 AAAAATTATTGAAATCTTT 489

RESULT 11
US-10-027-632-294129/c
Sequence 294129, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108927.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 294129
LENGTH: 627
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(627)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-294129

Query Match
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 AAAAATTATTGAAATCTTT 1153
DB 508 AAAAATTATTGAAATCTTT 489

RESULT 12
US-09-815-242-9987/c
Sequence 9987, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9987
LENGTH: 768
TYPE: DNA
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(768)
US-09-815-242-9987

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1249 ACCAGATGAGTTTCGATT 1268
DB 332 ACCAGATGAGTTTCGATT 313

RESULT 13
US-10-029-386-14103/c
Sequence 14103, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: AEOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14103
LENGTH: 199
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL135926.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EST HUMAN HIT: AU152688.1, EVALUATE 1.00e-06
OTHER INFORMATION: SWISSPROT HIT: Q09332, EVALUATE 7.30e+00
OTHER INFORMATION: NT HIT: U49016.1, EVALUATE 2.00e-10
US-10-029-386-14103

```
Query Match      1.1%; Score 19; DB 13; Length 199;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1259 GTTCTGATTTGGATTTT 1277
Db 63 GTTCTGATTTGGATTTT 45

RESULT 14
US-10-242-535A-18166/c
; Sequence 18166, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18166
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-242-535A-18166

Query Match      1.1%; Score 19; DB 12; Length 281;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 TGTAGCCTTCTTTCTGG 720
Db 64 TGTAGCCTTCTTTCTGG 46

RESULT 15
US-09-983-965-5678
; Sequence 5678, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5678
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Bos taurus
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: 55-LIB34-044-Q1-E1-F4
US-09-983-965-5678

Query Match      1.1%; Score 19; DB 10; Length 377;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1516 ATTCTCACTTGTTCCTG 1534
Db 158 ATTCTCACTTGTTCCTG 176

Search completed: February 2, 2004, 08:33:18
Job time : 1258.7 secs
```